MEDICAL IMAGE SEGMENTATION BY DEEP CONVOLUTIONAL NEURAL NETWORKS

Qingbo Kang

A thesis
in
The Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Applied Science (Software Engineering)
Concordia University
Montréal, Québec, Canada

August 2019
© Qingbo Kang, 2019

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Qingbo Kang**

Entitled: **Medical Image Segmentation by Deep Convolutional Neural Networks**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

_____ Chair
Dr. Denis Pankratov

_____ Examiner
Dr. Tien Dai Bui

_____ Examiner
Dr. Adam Krzyzak

_____ Supervisor
Dr. Thomas Fevens

Approved _____
Chair of Department or Graduate Program Director

_____ 2019 _____ _____
Dr. Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Medical Image Segmentation by Deep Convolutional Neural Networks

Qingbo Kang

Medical image segmentation is a fundamental and critical step for medical image analysis. Due to the complexity and diversity of medical images, the segmentation of medical images continues to be a challenging problem. Recently, deep learning techniques, especially Convolution Neural Networks (CNNs) have received extensive research and achieve great success in many vision tasks. Specifically, with the advent of Fully Convolutional Networks (FCNs), automatic medical image segmentation based on FCNs is a promising research field. This thesis focuses on two medical image segmentation tasks: lung segmentation in chest X-ray images and nuclei segmentation in histopathological images.

For the lung segmentation task, we investigate several FCNs that have been successful in semantic and medical image segmentation. We evaluate the performance of these different FCNs on three publicly available chest X-ray image datasets.

For the nuclei segmentation task, since the challenges of this task are difficulty in segmenting the small, overlapping and touching nuclei, and limited ability of generalization to nuclei in different organs and tissue types, we propose a novel nuclei segmentation approach based on a two-stage learning framework and Deep Layer Aggregation (DLA). We convert the original binary segmentation task into a two-step task by adding nuclei-boundary prediction (3-classes) as an intermediate step. To solve our two-step task, we design a two-stage learning framework by stacking two U-Nets. The first stage estimates nuclei and their coarse boundaries while the second stage outputs the final fine-grained segmentation map. Furthermore, we also extend the U-Nets with DLA by iteratively merging features across different levels. We evaluate our proposed method on two public diverse nuclei datasets. The experimental results show that our proposed approach outperforms many standard segmentation architectures and recently proposed nuclei segmentation methods, and can be easily generalized across different cell types in various organs.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Dr. Thomas Fevens who guides me into the deep learning and medical image processing field. Dr. Fevens is a very nice person and always inspires me to apply deep learning techniques on medical imaging problems. I really appreciate his support during the period of master study. Secondly, I want to appreciate those people who helped me with my study, research and thesis writing, including every members in our lab, thanks for their helpful discussion and corporation. Next, I would like to thank the Surgical Innovation program, the cross-disciplinary graduate program from McGill, ETS and Concordia. It provides me an unique opportunity to learn about the medical field and working with talented people from other fields: medical, business and engineering. It also provides me generous financial support and I can not finish my study in Montreal without that. Last but not the least, I want to thank my parents, for their selfless love and continuous encouragement.

Contributions of Authors

A Two-Stage Learning Framework for Nuclei Segmentation

Qingbo Kang: Model architecture design, coding, models training, experimental work, data analysis, writing, editing and proofing.

Qicheng Lao: Color normalization, model architecture design, editing and proofing.

Thomas Fevens: Research supervisor, funding, editing and proofing.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the first chapter, I will give a brief introduction of my thesis. First of all, I will describe the medical image segmentation problem. Secondly, how the deep learning techniques used for medical image segmentation will be discussed. Thirdly, the contributions of this thesis will be mentioned and finally, I give the outline of this thesis.

1.1 Medical Image Segmentation

Medical imaging techniques play a prominent role and have been widely used for the detection, diagnosis, and treatment of diseases [14]. There are many medical imaging modalities including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, positron emission tomography (PET) and so on. A typology of common medical imaging modalities used for different parts of human body which are generated in radiology is shown in Fig. 1.

Since this thesis focus on X-ray images and pathological images, we provide some details about these two kinds of imaging techniques in the following.

X-ray Images Since the German physicist Roentgen discovered X-rays in 1895, X-ray images have been used for clinical diagnosis for more than 100 years. Medical X-ray images are electron density metric images of different tissues and organs in human body. X-ray based imaging including 2D computer radiography, digital X-ray photography, digital subtraction angiography, mammography and 3D spiral computed tomography, etc., have been widely used in orthopedics [129], lungs , breast and cardiovascular [106] and other clinical disease detection and diagnosis. However, 2D X-ray images can not provide three-dimensional information of human tissues and organs. The automatic identification for 2D X-ray images is also difficult since there are overlaps in tissues and organs.

Pathological Images Pathological images refer to cutting a certain size of diseased tissue, using hematoxylin and eosin (H&E) or other staining methods to make the sliced tissue into a pathological slide, and then utilizing microscopic imaging techniques for cells and glands. By analyzing the pathological images, the causes, pathogenesis of the lesions can be explored to make a pathological

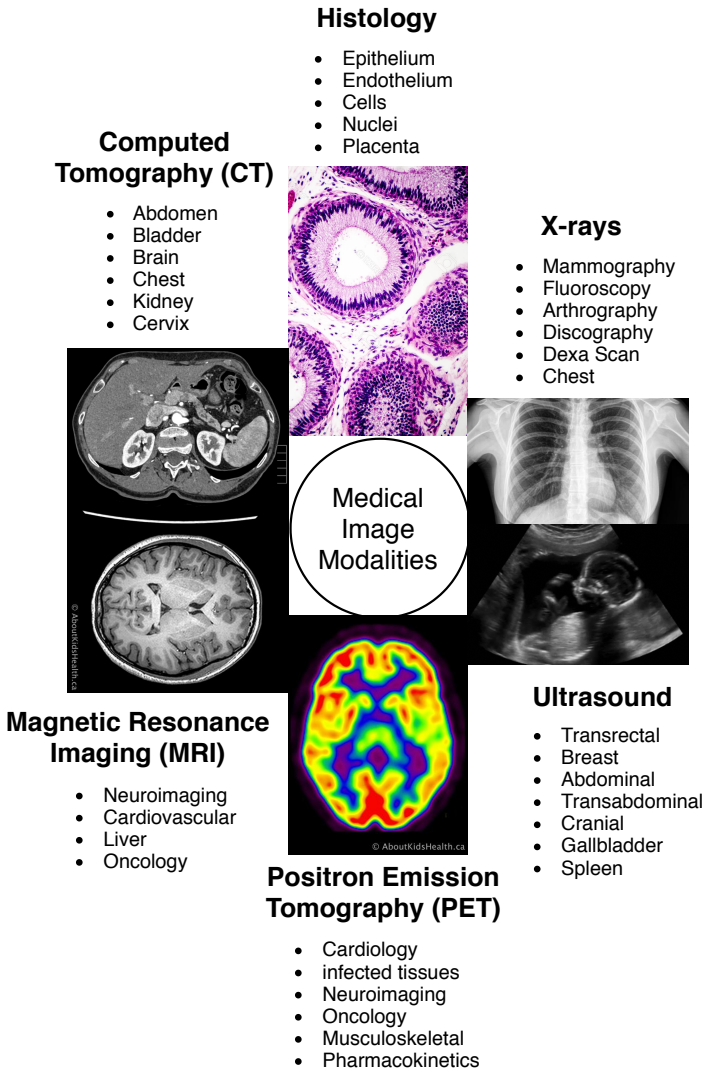

Figure 1: Typology of medical imaging modalities. Image is from [118].

diagnosis. Recently, with the advent of whole-slide imaging (WSI), it can obtain tumor spatial information such as nuclear direction, texture, shape, and structure and allows quantitative analysis of sliced tissue. A prerequisite for identifying these quantitative features is the need of detection and segmentation of histological primitives such as nuclei and glands [99].

Medical image segmentation is a complex and critical step in medical image processing and analysis. The purpose of medical image segmentation is dividing an image into multiple non-overlapping regions based on some criterion or rules such as similar gray level, color, texture etc.. Based on various traditional techniques, many researchers proposed a great number of automated segmentation approaches such as thresholding, edge detection, active contours and so on [115, 123]. After that, machine learning based methods have dominate this field for a long period. Machine learning rely on hand-crafted features, therefore how to design suitable features in different field and different

imaging modalities has become a primary concern and a key factor for the success of such a segmentation system. However, due to the complexity and diversity of medical images, the segmentation of medical images continues to be a challenging problem.

1.2 Deep Learning for Medical Image Segmentation

Deep learning has been widely used and achieves great success in many areas such as computer vision, speech analysis and natural language processing [83]. In contrast to traditional machine learning techniques which based on hand-craft features for different task, deep learning directly learns representation features from huge amount of data. Specific to medical image segmentation field, deep leaning techniques based approaches especially approaches based on Convolution Neural Networks (CNNs) have received extensive attention and research, many works have been proposed and achieved superior performance compared to segmentation methods based on other techniques [94, 124]. Many CNNs based segmentation network such as FCN [95], U-Net [120], V-Net [102] and their variants or improvements [28, 39, 162, 76, 110, 3, 114, 51, 50] have been proposed and achieve state-of-the-art performance on numerous medical image segmentation tasks.

1.3 Contributions of this Thesis

In this thesis, we focus on two medical image segmentation tasks, lung segmentation in chest X-ray images and nuclei segmentation in histopathological images.

For the lung segmentation problem, we apply FCN and U-Net, the two most widely used segmentation model for medical image segmentation, on this task. We evaluate the performance of these models on three publicly available chest X-ray datasets, the experimental results demonstrate the superior performance of deep learning based segmentation models.

For the nuclei segmentation problem, we propose a novel nuclei segmentation approach based on a two-stage learning framework and Deep Layer Aggregation (DLA) [156]. We convert the original binary segmentation task into a two-step task by adding nuclei-boundary prediction (3-classes) as an intermediate step. To solve our two-step task, we design a two-stage learning framework by stacking two U-Nets. The first stage estimates nuclei and their coarse boundaries while the second stage outputs the final fine-grained segmentation map. Furthermore, we also extend the U-Nets with DLA by iteratively merging features across different levels. We evaluate our proposed method on two public diverse nuclei datasets. The experimental results show that our proposed approach outperforms many standard segmentation architectures and recently proposed nuclei segmentation methods, and can be easily generalized across different cell types in various organs.

1.4 Outline of this Thesis

This thesis is organized as follows: Chapter 2 will reviews some related works of this thesis, specifically the literature reviews for the two segmentation tasks, some details of CNN and two segmentation model: FCN and U-Net. Chapter 3 will presents our lung segmentation work in chest X-ray images and the associated experimental results. Chapter 4 will presents our nuclei segmentation work and the corresponding experimental results. In Chapter 5, we will conclude this thesis and discuss some future work and research directions.

Chapter 2

Related Works

This chapter will cover related works of this thesis. Specifically I will briefly review some literature for the two segmentation tasks which this thesis focuses on, i.e. lung segmentation in chest X-rays and nuclei segmentation in histopathological images. After literature reviews, CNNs and the Fully Convolutional Neural Networks will be described in this chapter.

2.1 Literature Reviews for Medical Image Segmentation

In this section, the existing approaches for the two medical image segmentation tasks will be respective reviewed.

2.1.1 Lung Segmentation in Chest X-rays

Over the past decades, researchers have proposed a number of methods to segment the lung field from chest X-ray images. These methods can be divided into four categories [141, 62]: rule-based segmentation [116, 149, 15, 18, 89], pixel classification-based segmentation [101, 55, 145, 138, 5, 142, 25], deformable model-based segmentation [67, 157, 4, 32, 140, 125, 52] and hybrid segmentation [142, 34, 17].

Rule-based Segmentation

The rule-based segmentation methods aim to obtain the expected target region of interest after image pixels are processed through a series of steps and rules. Most of the early proposed lung field segmentation algorithms fall into this category [116, 149, 15]. Some techniques like threshold segmentation, region growth, edge detection, ridge detection, mathematical morphology, geometric models, and so on are used to find the edge of the lung area based on the characteristics of lung structure [18, 89].

Pixel Classification-based Segmentation

A series of feature vectors are calculated for each pixel in the image, and some pattern recognition techniques are used to mark the category of each pixel belongs to according to the feature vector [101]. For the digital X-ray chest radiology segmentation problem, the pixel classification method is to assign each pixel in the chest radiograph image with the corresponding anatomical structure (such as lung and background, or heart, mediastinum and diaphragm, etc.) through a classifier. The classifier can use pixel point gray information, spatial position information, texture statistics information, etc. as feature vectors then obtain the labels through training of neural network [55, 138, 5], K nearest neighbor (KNN) classifier [142], support vector machine (SVM) [25], Markov random field model [145], etc.

Deformable Model-based Segmentation

The segmentation method based on deformable model belongs to the top-down strategy. Firstly, an overall model for understanding the target is generated according to the content of the image, and then the image feature is applied to fit the model to the best match and the target object is segmented. After more than 20 years of research, from elastic model, active contour model [67, 157, 4] to active shape model [32, 140, 125, 52], the deformable model has been developed and widely used in the field of image segmentation. In the field of lung segmentation, active contour models and active shape models have received the most attention from researchers. Iglesias *et al.* [67] first use the active contour model with shape constraints for lung segmentation, and studied the effects on the segmentation results of different parameters in the active contour model. Yu *et al.* [157] propose a lung segmentation method based on shape regularized active contour. Annangi *et al.* [4] present a work by using level set energy to segment the lungs from chest X-rays. Cootes *et al.* [32] propose active shape models. Van *et al.* [140] present a segmentation method based on active shape models with optimal features. Shi *et al.* [125] use the active shape model based on scale-invariant feature transform (SIFT) features to segment the lungs. Guo and Fei [52] develop a minimal path searching method for active shape model based segmentation for chest X-rays.

Hybrid Segmentation

Combining multiple segmentation methods and overcoming shortcomings of one method by another method. It is hoped that the combined use of multiple methods can complement each other and make the segmentation result better. After using the Active Shape Model (ASM), Active Appearance Model (AAM), and Pixel Classification (PC) to segment the lungs, Van Ginneken *et al.* [142] proposed a joint ASM, AMM, and PC method to segment images. In order to obtain the independent segmentation results of the pixels of ASM, AMM, and PC, each pixel is voted by using the classification results of the three methods, and each pixel is classified according to the majority principle. Another strategy is utilizing the segmentation result of one method as the input of another method for the second segmentation, such as ASM/PC, PC/ASM, and so on. Candemir *et al.* [17] present a hybrid method based on nonrigid registration and anatomical atlas as a guide combined

with graph cuts for refinement.

2.1.2 Nuclei Segmentation in Histopathological Images

Nuclei segmentation has been studied for decades and a large number of methods have been proposed [147]. Most of the traditional nuclei segmentation methods based on these following algorithms: intensity thresholding (such as OTSU [112, 150, 122]), image morphological operations [160], watershed transform algorithm [151, 144], active contours [73, 111, 104, 29, 19, 100, 153, 20, 148, 21, 143, 54, 35, 137], clustering (such as K-means [98]), graph-based segmentation methods [42], supervised classification and their variants or combinations.

Intensity Thresholding

The most basic and simplest algorithm for nuclei segmentation may be intensity thresholding. Using a global threshold value or some locally adaptive threshold values to convert the input image to a binary image is widely used in image processing field. The method for choosing the specific thresholding value is related to the task and the input image. Specific for nuclei segmentation, the intensity distribution of pixel values between nuclei (foreground) and the background is persistently distinct. One of the most famous locally adaptive algorithms is OTSU [112], which selects a threshold by maximizing the variance between the foreground and the background. In order to tackle the problem of non-consistent intensity values within an image, an extension of this method is to divide the full image into numerous sub-images and perform thresholding individually [122], but it requires additional parameters thus can't perform automatically.

Image Morphological Operations

Mathematical morphology is one of the most widely used techniques in image processing field. The basic operations including erosion, dilation, opening and closing. For nuclei segmentation task, morphological operations often cooperate with other methods to achieve better segmentation performance. For example, [160] presents an unsupervised nuclei segmentation method which using morphology to enhance the gray level values of the nuclei.

Watershed

Watershed transform is one of the most important image segmentation algorithms. It can be classified as a region-based segmentation method which utilizing a region growing strategy, specifically, it starts with some seed points and then iteratively adds image pixels which satisfies some requirements to regions. [151] proposed a marker-controlled watershed to avoid over-segmentation problem in segmenting clustered nuclei. [144] utilized marker-controlled watershed segmentation for nuclei segmentation in H&E stained breast biopsy images.

Active Contours

Active contour models or deformable models are extensively studied and used for nuclei segmentation. With some initial starting points, an active contour evolves toward the boundaries of desired region or objects by minimizing an energy functional. The energy of the active contour model (also known as Snake) is formulated as a linear combination of three terms [73]: internal energy, image energy and constraint energy. The internal energy controls the smoothness and continuity of the contour, the image energy encourages the Snake to move toward features of interest and the constraint energy can be based on the specific object. The two major implementations of active contour models for nuclei segmentation are geodesic snakes (or level set models) which are with implicit contour representations and parametric snakes which are with explicit contour representations [147]. A contour is implicitly represented as the zero level set of a high-dimensional manifold in a geodesic model [35, 111]. There are mainly two types geodesic models: edge-based level set models [19, 100, 153, 20] which rely on the image gradient to terminate contour evolution and region-based level set models [21, 143] which based on the Mumford-Shah functional [104]. Region-based models are more robust to noise and weak edges compared with edge-based models [147]. Han *et al.* [54] present a topology preserving level set model to preserve the topology of the implicit curves or surfaces throughout the deformation process. Taheria *et al.* [137] propose a nuclei segmentation approach which utilizes a statistical level set approach along with topology preserving criteria to evolve the nuclei border curves. While in a parametric active contour model, a continuous parameter is explicitly used to represent a contour. The traditional Snake model [73] moves contours toward desired image edges while preserving them smooth by searching for a balance between the internal and external force. A balloon snake [29] is formed by introducing a pressure force to increase the capture range of the external force. On the other hand, [148] replaced the external force with a gradient vector flow (GVF) to handle the problems of poor convergence to boundary concavities and sensitive initialization.

Clustering

Clustering is the process of dividing a collection of data objects into multiple subsets. Each subset is called a cluster. Clustering makes the objects in the cluster have high similarity, but it is not very similar to objects in other clusters. Different clustering algorithms may produce different clusters on the same dataset. Cluster analysis is used to gain insight into the distribution of data, observe the characteristics of each cluster, and further analyze the characteristics of specific clusters. Since a cluster is a subset of data objects, the objects in the cluster are similar to each other and not similar to the objects in other clusters. Therefore, the cluster can be regarded as a "recessive" classification of the dataset, and cluster analysis may find the unknown subset of the dataset. Clustering is unsupervised learning, unsupervised learning refers to the search for implicit structural information in unlabeled data. For nuclei segmentation task, clustering is usually used as an intermediate step such as extract object boundary. Popular clustering algorithms including K-means [98], Fuzzy c-means [10] and EM algorithm [36]. [78] presented a K-means clustering based approach for nuclei segmentation in H&E and immunohistochemistry (IHC) stained pathology images. [6] designed a nuclei segmentation method based on manifold learning which utilizing K-means to segment nuclei

and nuclei clumps. [16] proposed a parallel Fuzzy c-means based approach for nuclei segmentation in large-scale images, which can be used to process image which has high resolution such as WSI.

Graph-based Methods

A graph-based image segmentation method [42] treats an image as a weighted graph. Each node in the graph represents a pixel or super-pixel in the image, and each edge weight between the nodes corresponds to the similarity of adjacent pixels or super-pixels. In this way, a graphic can be divided into multiple regions according to a criterion, each region represents an object in the image. Typical example graph-based methods including Max-Flow/Min-Cut algorithms [49, 13, 12], normalized cut [146] and Conditional Random Fields (CRF) [82]. The Max-Flow/Min-Cut algorithms solve the image segmentation problem by minimizing an energy function. While the normalized-cut algorithm attempts to divide the set of vertices of an undirected graph into multiple disjoint classes, so that the similarity between classes is very low, the similarity within the class is very high, and the size of the class should be as balanced as possible. CRF formulates segmentation task as a pixel-wise classification or labelling task and assigns the labels of each pixel or super-pixel based on the observations, this method can be classified as a discriminative graphical model.

Supervised Classification

A number of nuclei segmentation methods based on supervised machine learning have also been proposed. There are two categories methods for this task, i.e. pixel-wise classification or superpixel-wise classification. For pixel-wise, the label of each pixel of determined by a learned model with some criteria. While for superpixel-wise classification, a set of candidate regions for nuclei are first segmented from the input by a learned model. The general pipeline of this method is first apply some feature extraction algorithms to extract image features from input image and then feed into classifiers such as K-NN, SVM [133], Bayesian, etc. [77] presents a supervised learning algorithm for nuclei segmentation in follicular lymphoma pathological images. The local Fourier transform features are firstly extracted from the image, then a K-NN classifier is applied to determine the label of each pixel.

2.2 Convolutional Neural Network

Deep learning [48] has been widely used and achieved notable success in many domains such as computer vision [79, 128, 119, 95, 45], natural language processing [30, 31, 154], speech recognition [60, 37, 161]. CNNs [86] are a special kind of feed-forward network with sparse connectivity and parameter sharing, which are particular designed for dealing with data that has grid-like topology such as image data [48]. CNNs have achieved remarkable performance in plenty of computer vision tasks including image classification [79, 128, 58, 59, 63], image segmentation [95, 109, 7, 56, 24], face recognition [113, 22], image style transfer [45, 96] etc.

CNNs are motivated by the mechanism of receptive field in biology. In 1959, David Hubel and Torsten Wiesel discovered that there are two types of cells in cat's primary visual cortex: simple

cells and complex cells. These two kinds of cells responsible for tasks in different levels of visual perception [64, 65, 66]. The receptive field of the simple cell is long and narrow, and each simple cell is only sensitive to the light with the specific orientation in the field, while the complex cell is aware of the light of an orientation in the field moving along a specific direction. Inspired by this observation, in 1980, Kunihiko Fukushima proposed a multi-layer neural network with convolution and sub-sampling operations: Neocognitron [44]. After that, Yann LeCun introduced the back-propagation (BP) algorithm into CNNs in 1989 [84] and achieved great success in handwritten digit recognition [85].

AlexNet [79] is the first modern deep CNN model, which can be considered as the beginning of a real breakthrough of deep learning techniques for image classification. AlexNet does not require pre-training and layer-wise training, on the contrary, it uses many techniques that are widely used in modern deep CNNs, such as parallel training using GPU, ReLU as a nonlinear activation function, dropout [61, 130] to prevent over-fitting, and data augmentation to improve the performance of the model, etc. These techniques have greatly promote the development of end-to-end deep learning models. There are many CNN models have been proposed after AlexNet, such as VGG [128], Inception v1 [135], v2 [136], v4 [134], ResNet [58, 59] DenseNet [63] and so on.

Currently, CNNs have become the dominating models in the field of computer vision. By introducing skip connection across layers, the depth of a CNN may beyond one thousand layers. However, no matter how deep a CNN model is, the basic building blocks of it stay the same. In general, it may consists of convolution layers, pooling layers and fully-connected layers. This section will give some details of each building blocks in a CNN model.

2.2.1 Artificial Neural Network

Artificial Neural Networks (ANNs) are artificial computational systems which were mainly motivated by biological neural systems in human brain. The most fundamental element in ANNs is neurons or nodes, a typical ANN consists of numerous neurons and weighted connections between these neurons. Neurons receive input signals from connections and perform some operations then generate outputs [70, 152]. Neurons are grouped by layer and ANNs may have multiple layers, the number of layers is called the depth of ANNs. An ANN can be trained to approximate a particular function by adjusting the weights of connection. According to the connection pattern, ANNs can be divided into feed-forward networks in which there is no loop or feedback connections, and recurrent networks in which there has feedback connections [48]. In particular, a feed-forward network with all neurons in current layer have connections with all neurons in the next layer is called fully-connected network. Fig. 2 shows an example of 3 layers (input layer, hidden layer and output layer) fully-connected feed-forward network.

2.2.2 Convolution Operation

Convolution operation initially is an important operation in mathematics, it also has a broad usage in signal and image processing. Since the convolution used in neural networks has some slightly differences compared to the convolution used in pure mathematics, the convolution described here is

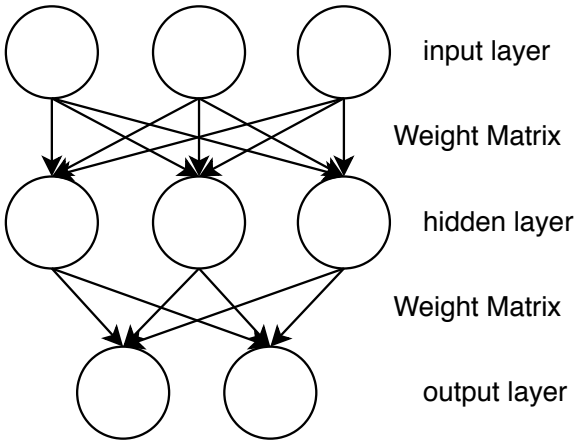

Figure 2: A typical fully-connected feed-forward neural network with depth 3.

just used in neural networks. When apply on imaged, the convolution usually has a two-dimensional discrete form. Formally, let I be an image and K be a kernel, the two-dimensional discrete convolution is:

$$P(i,j) = (I * K)(i,j) = \sum_{u=0}^{i} \sum_{v=0}^{j} I(u,v)K(i-u, j-v) \tag{1}$$

where the range of u is $[0, i)$ and the range of v is $[0, j)$, i and j are the width and height of the kernel, respectively. In the context of deep learning and image processing, the main function of convolution is to obtain a new set of features or representations by sliding a convolution kernel (i.e. filter) on an image. In practice, many deep learning libraries such as TensorFlow [1], Theano [9] and Caffe [72] use cross-correlation operation instead of convolution operation, which can reduce unnecessary computation cost significantly. Given an image I and kernel K, the cross-correlation is defined as:

$$P(i,j) = (I * K)(i,j) = \sum_{u=0}^{i} \sum_{v=0}^{j} I(i+u, j+v)K(u,v) \tag{2}$$

For the purpose of feature extraction, convolution and cross-correlation are equivalent, the only difference between convolution and cross-correlation is whether the kernel is flipped. Fig. 3 shows an example of 2-D convolution operation without kernel flipping.

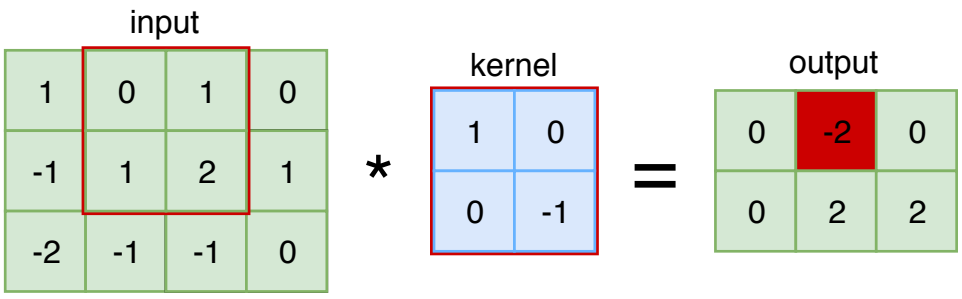

Figure 3: An example of 2-D convolution operation without kernel flipping. The output in the red square is the convolution result of the red squared input region and the kernel.

2.2.3 Local Connectivity and Parameter Sharing

Compared to ordinary neural network layers, convolution layer has two important properties: local connectivity and parameter sharing.

Local Connectivity

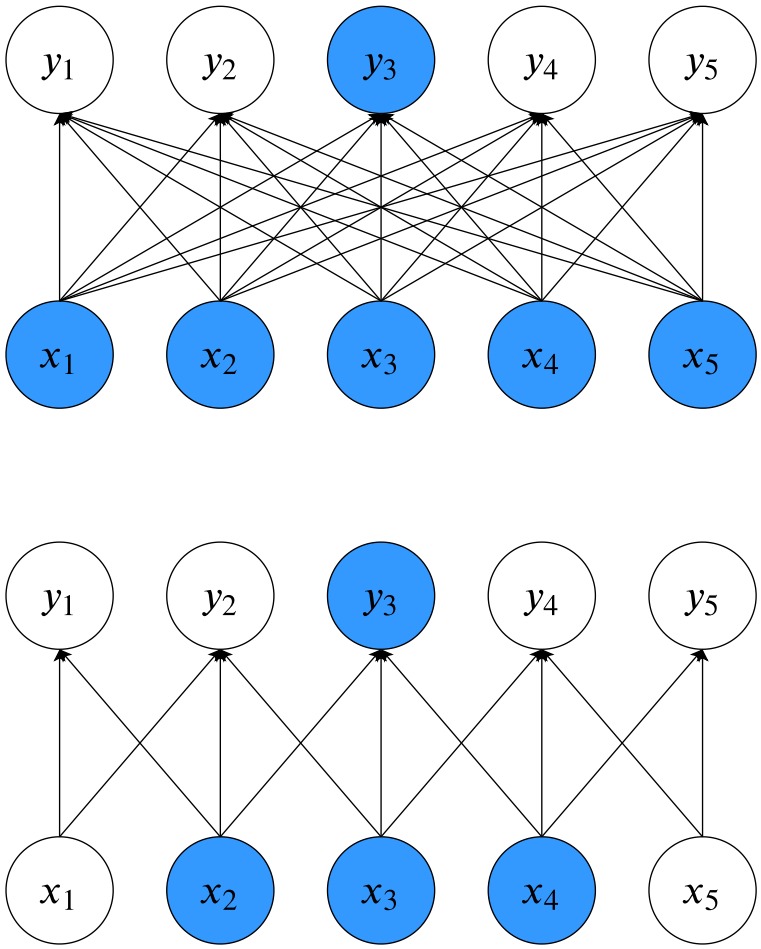

Figure 4: Schematic diagram of local connectivity. The upper half is fully connected layer and the bottom half is locally connected layer. Image is from [48].

All the neurons in one convolution layer only connect with neurons in a small local region of previous layer. The local region is called the receptive filed of this neuron. Local connectivity (also known as sparse connectivity, sparse weights and sparse interactions) ensures that the learned filter has the strongest response to local input features and also can decrease the number of parameters of a CNN model dramatically. Fig. 4 schematically illustrates the local connectivity property. More precise, the upper half of Fig. 4 shows the connectivity pattern of a fully connected layer while the bottom half describes the local connectivity pattern of a convolution layer. In the upper half, the above row is the matrix multiplication result with fully connectivity, the blue circles in the bottom

row affect the result output y_3 and are called the receptive field of y_3. Since it's fully connected, all the inputs affect y_3. While in the bottom half, the above row is the convolution result of kernel with width 3 applies on the bottom row. With local connectivity, only 3 inputs affect y_3.

Parameter Sharing

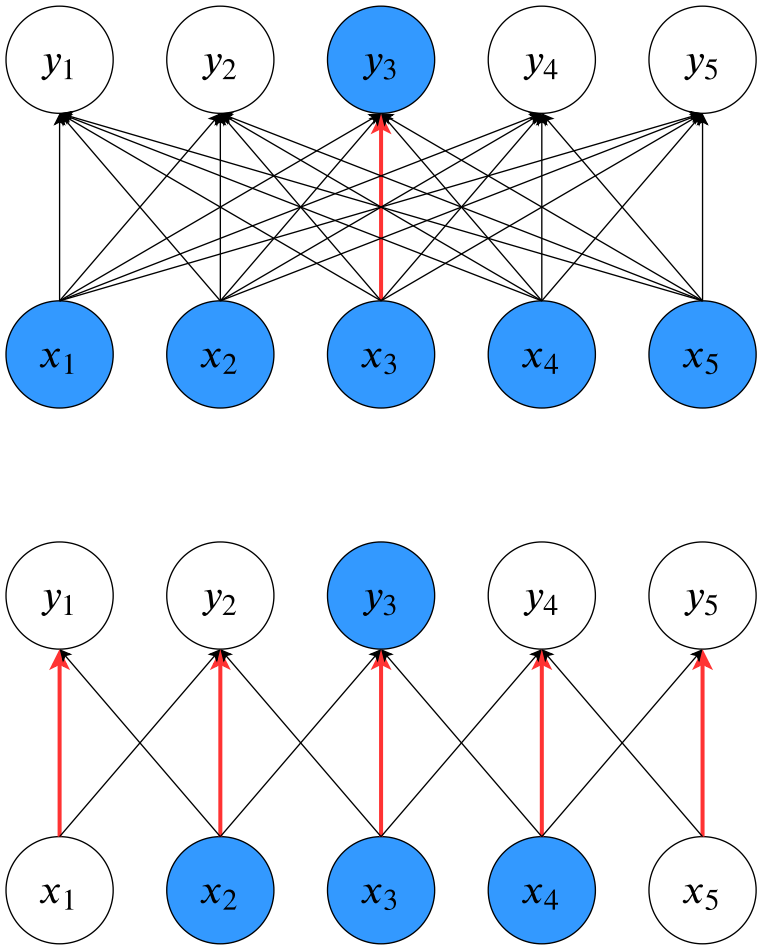

Figure 5: Schematic diagram of parameter sharing. The upper half is without parameter sharing and the bottom half is with parameter sharing. Image is from [48].

In a CNN, the parameters are the same for a convolution operation applies for every neurons in one layer. This means for one layer, we do not need learning separate sets of weights for every location, we just need learning one set of weights and then applying them everywhere. This property further reduces the number of parameters. Fig. 5 demonstrates the parameter sharing property. The red arrows in Fig. 5 represent the connections that use an unique parameter in two different situations. In the upper half, the situation without parameter sharing, the parameter is unique and used only once. While in the bottom half, the parameter of the central element of a convolution of kernel with width 3 is used at all input locations because of parameter sharing.

2.2.4 Activation Function

The main purpose of an activation function in a neural network is to provide nonlinear modeling ability for the neural network. A neural network without nonlinear activation function can only express linear mapping, and no matter how many layers this network has, it is equivalent to one single-layer neural network. In general, neurons receive some input signals, perform some operations or functions such as weighted sum, and optionally followed by nonlinear activation functions. Typical activation functions including Sigmoid, tanh, Rectified Linear Units (ReLU) [105] and Leaky-ReLU [97]. Fig. 6 gives the corresponding formula and figure of some widely used activation functions in neural networks.

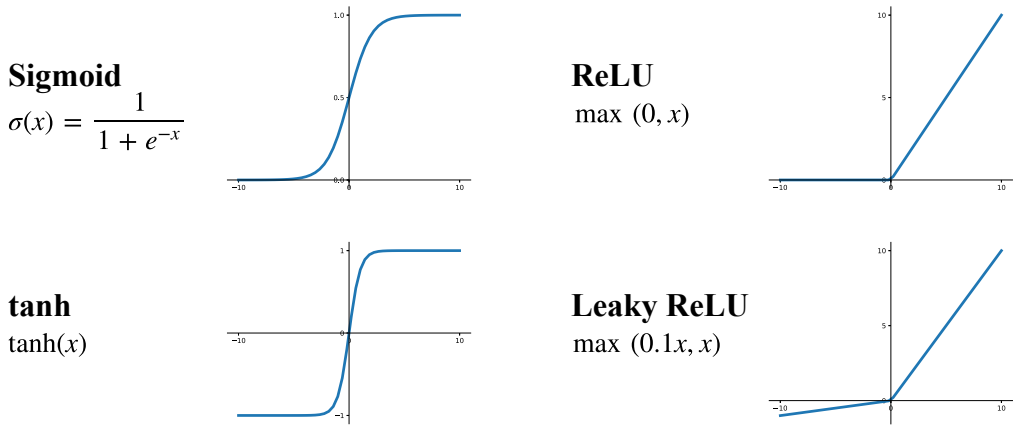

Figure 6: Some widely used activation functions in neural networks.

The Sigmoid function is the most widely used non-linear activation function historically, it converts the continuous real-valued input to an output between 0 and 1 and particularly suitable for classification problems. But in recent years, it has fallen out of favor and rarely ever used since it has three major drawbacks. The first drawback is that it can saturate and kill gradients and cause gradient exploding/vanishing problem [47]. The second drawback of Sigmoid is the outputs is not zero-centered and this will slow down the convergence of deep neural networks. The last drawback is that the Sigmoid has a power operation which is a relatively time-consuming operation and will increase the training time for deep networks. The tanh function solves the second drawback of Sigmoid, i.e. the not zero-centered problem, but the gradient exploding/vanishing and the power operation problems still exist. The ReLU solves the gradient exploding/vanishing problem in positive interval and is computational efficient but the outputs of ReLU is not zero-centered. It also has dead ReLU problem which means some neurons may never be activated (the corresponding parameters never be updated). However, the ReLU function is still the most commonly used activation function nowadays [79]. In order to tackle the dead ReLU problem, the Leaky ReLU function was proposed [57] which has a small negative slope. In theory, the Leaky ReLU is better than ReLU since there will be no dead ReLU problem, but in practice, it does not fully prove that the Leaky ReLU is always better than ReLU.

There are a great number of activation functions used in neural networks and each of them has different properties, how to choose the right activation function is depending on the specific task you are perform (i.e. the function that you are trying to approximate). In addition, different activation functions can be used in different layers in one CNN architecture, for example, many deep CNNs for image classification use ReLU as activation function in hidden layers and use Sigmoid as activation function in output layer [79, 128, 135].

2.2.5 Pooling

Pooling layer, also known as sub-sampling layer, it performs down sampling operation on the feature maps thus decreasing the dimension of feature maps and thereby reducing the number of parameters. Since pooling operation summaries some statistics of the neighboring outputs in previous layer, it enables the feature representations after pooling operation approximately unchanging to small translation. The size of the pooling layer is the window size which used for calculation, and the stride of the pooling layer is the number of pixels between every calculation. There are two commonly used pooling functions: max-pooling which choose the maximum value and average-pooling, in contrast, selects the average value. Fig. 7 illustrates a max-pooling operation with size 2×2 and stride 2×2.

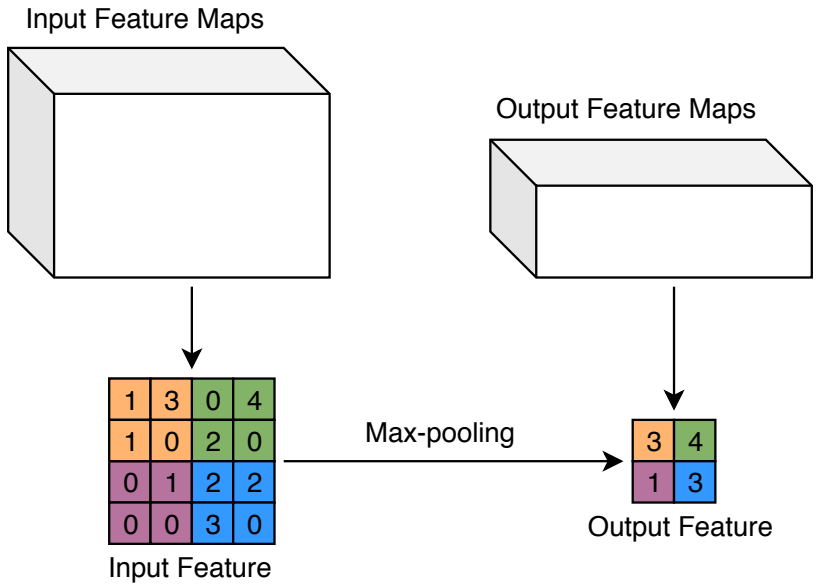

Figure 7: Max-pooling operation (size 2×2 and stride 2×2).

2.2.6 Typical CNN Structure

For a classification task, the architecture of a typical CNN is composed of a stack of convolution layers, pooling layers and fully-connected layers. At present, the pattern of most widely used CNN structure is shown in Fig. 8. A convolution layer usually involves a convolution operation followed by an activation function. A convolution block consists of successive M convolution layers and b

pooling layers. N consecutive convolution blocks can be stacked in a CNN, finally followed by K fully-connected layers.

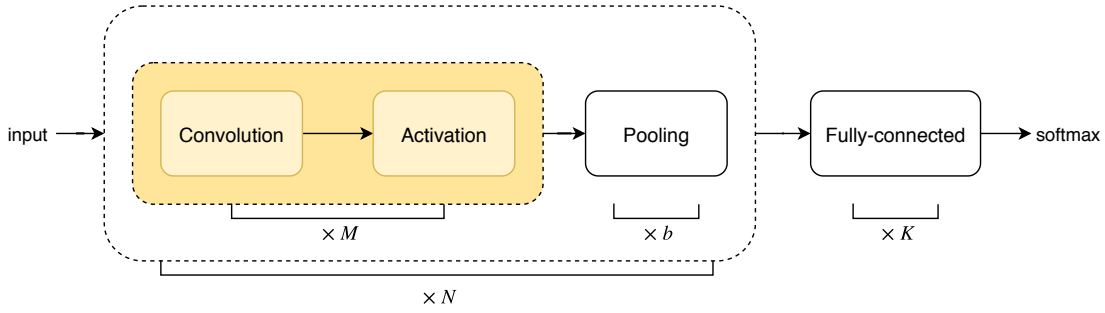

Figure 8: A typical CNN structure for a classification task.

The purpose of fully-connected layers in a classification task is to map the result features of the convolution layers and pooling layers to class labels. Clearly, since the fully-connected layers have a huge number of parameters, they are wasteful and cost a large amount of computational power, thus cannot scale to large input image. In some CNN architectures such as GoogLeNet [135], there is no fully-connected layers.

2.2.7 Training Neural Networks

Training of a neural network means solving one particular case of optimization problem: finding the parameters or the weights of the connections θ in a neural network that minimize a predefined loss function $\mathcal{L}(Y, f(X, \theta))$. The loss function measures the performance of the neural network on the data , specifically, it evaluates the degree of inconsistency between the predicted labels $f(X, \theta)$of the model given the current weights θ of the model and the ground truth labels Y. Two commonly used loss functions are Mean Squared Error (MSE) and Cross-Entropy (CE). MSE calculates the average of the squared differences between the predicted labels and the true labels:

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2 \tag{3}$$

where m is the number of data samples and $\hat{Y}_i$ is the i-th predicted label. MSE is usually used in regression problem, while CE is often used in classification problem, for a binary classification problem (i.e. $Y_i = \{0, 1\}$), the CE can be defined as:

$$\text{CE} = -\frac{1}{m}\sum_{i=1}^{m}(Y_i \log \hat{Y}_i + (1 - Y_i)\log(1 - \hat{Y}_i)) \tag{4}$$

where Y_i and $\hat{Y}_i$ are the ground truth labels and predicted labels, respectively. In the context of machine learning, gradient based learning algorithms are widely used to train neural networks. Specifically, BP algorithm [84] is used to compute the gradients for each parameter based on the total loss value of the model. The core idea of BP is utilizing chain rule repeatedly to calculate partial derivatives for each parameter in the model. Basically, it starts from the last layer, calculates

the error vector in reverse, continuously applies the chain rule to calculate the loss value of the cumulative gradient inversely, thus minimizes the loss function. After obtain all the gradients, gradient based learning algorithms is generally used to update the parameters. Specifically, gradient descent technique add an appropriate negative gradient on the original parameter:

$$\theta_{n+1} = \theta_n + \Delta\theta(n)$$
$$\Delta\theta(n) = -\alpha\frac{\partial\mathcal{L}}{\partial\theta(n)}$$

where α is called learning rate, $\mathcal{L}$ is the total loss value and n is the iteration number. When just part of the examples (mini-batch) from the training set are used for loss function calculation, the algorithm used for updating parameters based on this loss value is known as stochastic gradient descent (SGD). Since different initialization strategies of the parameters and the use of only partial samples during the parameter update process, SGD can only find the local optimal solution.

Based on the gradient descent learning algorithm, many optimization algorithms for training neural network are proposed. Momentum [117] is a method designed for accelerating learning process of SGD by introducing a hyper-parameter called momentum which is derives from physical analogy. Recently, many adaptive learning rate based optimization methods have been introduced, such as AdaGrad [40], Adam [75] and AdaDelta [158].

The training process can be divided into two categories, online learning and batch learning. Online learning usually selects one data sample randomly from the training set then learn one by one. The main advantage of online learning is small computational cost, however it converges pretty slow. Batch learning utilizes all data samples in the training set which benefits the loss calculation based on all data, but the computation is huge and only suits for the situation when has very small data samples. In practice, the most widely used training strategy is mini-batch learning which is a trade-off between the above mentioned two categories. Generally, the traversing of the entire training set of learning process is defined as one epoch.

2.2.8 The Initialization of Parameters

The initialization strategy of parameters in a CNN model has a big influence on the convergence speed and the performance of the model. Next, three most widely used parameters initialization methods will be described.

Gaussian Initialization

In this initialization, parameters are initialized with random values which selected from a specified Gaussian distribution $N(\mu, \sigma^2)$, the mean value and the variance of the Gaussian distribution are pre-defined and fixed.

Xavier Initialization

Xavier initialization was proposed by Glorot and Bengio [47], the initial variance of Gaussian distribution is no longer pre-defined and fixed but determined by the input layer of the current layer and

the number of neurons in the input layer. Suppose the number of neurons in the input layer is n_{in}, and the number of neurons in the output layer is n_{out}. The initial variance is:

$$Var = \frac{2}{n_{in} + n_{out}} \tag{5}$$

then a Gaussian distribution with zero mean and Var variance is used for parameters initialization.

MSRA Initialization

MSRA initialization was presented by He *et al.* [57]. Unlike the Xavier initialization, MSRA initialization uses different initial variance for Gaussian distribution to obtain a much more robust initialization. The initial variance of MSRA is:

$$\begin{aligned} Var &= \sqrt{\frac{2}{n_l}} \\ n_l &= k_l^2 d_{l-1} \end{aligned} \tag{6}$$

where k_l is the kernel size of convolution, and d_{l-1} is the number of convolution kernel in $(l-1)$-th layer.

In conclusion, Xavier initialization is more suitable for the network which uses Sigmoid as activation while MSRA initialization works better for the network uses ReLU as activation.

2.2.9 Batch Normalization

Training deep neural networks including deep CNNs is extremely challenging, one of the most important reasons is that deep neural networks may consist of a large number of layers, and the parameters of all layers are updated simultaneously. Every parameter update in one layer will change the input data distribution of all subsequent layers, even a small change in low layers' data distribution will cause exponentially change of high layers' data distribution. In order to train the model, we need to be very careful to set the learning rate, parameters initialization method and parameter update strategy. This kind of data distribution change in different layers is called the internal covariate shift [68]. In order to solve this problem, Ioffe and Szegedy [68] proposed an approach called batch normalization. Basically, batch normalization is an adaptive reparametrization approach which is aiming for making the training of deep neural networks easier. The details of batch normalization will be described in the following.

For a mini-batch with size m, it has m activation values which can be denoted as $\mathcal{B} = \{x_{1...m}\}$. Firstly, the mean and variance of the batch are computed:

$$\begin{aligned} \mu_{\mathcal{B}} &= \frac{1}{m}\sum_{i=1}^{m} x_i \\ \sigma_{\mathcal{B}}^2 &= \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \end{aligned} \tag{7}$$

where $\mu_{\mathcal{B}}$ is the mean value and $\sigma_{\mathcal{B}}^2$ is the variance of the mini-batch. After that, the normalized activation values $\hat{x}_{1...m}$ are:

$$\hat{x}_{1...m} = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \tag{8}$$

where ϵ is a small constant which used to avoid division by 0. Finally, the normalized activation values can be obtained by:

$$y_i = \gamma \hat{x}_i + \beta \tag{9}$$

where γ and β are learned parameters that allow the normalized activation values to have any mean and standard deviation. Batch normalization can apply on any types of layer in neural networks, [68] places the batch normalization layer before the activation function,

$$z = g(\mathrm{BN}(Wx)) \tag{10}$$

where BN stands for batch normalization and $g(*)$ is the activation function.

2.3 Fully Convolutional Neural Networks

For image segmentation task, since the proposal of FCN [95], which FCN stands for Fully Convolutional Networks, it attracts active research and many works based on FCN have been proposed [109, 120, 136, 102, 23, 7, 92]. Considering the output of image segmentation is a pixel-wise classification map instead of one single class label for image classification, the main idea of FCN is replacing fully-connected layers in a classification network with convolution layers thus make the network fully convolutional. Furthermore, U-Net [120] is an architecture based on FCN and has been widely proven to have superior performance for medical image segmentation. These two networks will be discussed in this section.

2.3.1 FCN

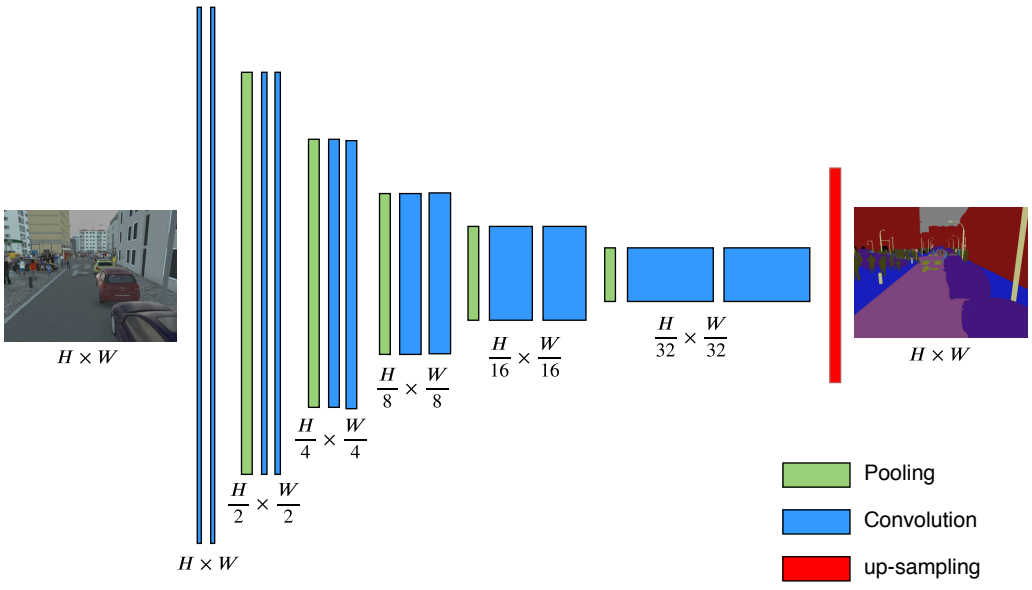

Figure 9: The FCN-32 network structure. Green box represents pooling operation, blue box represents convolution and activation operation and red box represents up-sampling operation.

As mentioned above, the main contribution of FCN is convolutionalization which means replacing fully-connected layers with convolution layers. There are various advantages of FCN compared to CNN with fully-connected layers. First of all, the FCN can process input image in different sizes, i.e. the resolution of the input image for a FCN is not fixed. Secondly, fully-connected layer usually has a huge amount of learnable parameters compared with convolution layer, and thus needs lots of memory to store the model and computation to train the model. The replacing of fully-connected layers with convolution layers in FCN make the whole network architecture fully convolutional.

Furthermore, FCN introduces up-sampling operation to recover the dimension of output feature maps back to original input dimension. In this way, a 2-dimensional feature map can be obtained, followed by a softmax function to generate a pixel-wise labelling map. Fig. 9 demonstrates the detailed structure of FCN32, in which the name FCN32 means it directly up-samples the features in the lowest resolution (32x up-sampling) back to the original resolution.

Transposed Convolution

FCN adopts transposed convolution (also known as deconvolution, backwards convolution) [159] to perform up-sampling. Although it is called transposed convolution, in fact, it's not the inverse operation of convolution. Transposed convolution is a special kind of forward convolution, it first enlarges the size of the input image by padding, then rotates the convolution kernel (matrix transpose) and performs forward convolution. The kernel weights of transposed convolution can be learned by backpropagation from the network loss. The transposed convolution enables the prediction of the segmentation network is pixel-wise, therefore make the learning of the whole network end-to-end.

Skip Layer

For the task of image segmentation, global information contains semantics of the whole image and local information indicates specific location of each object. In order to obtain accurate segmentation map, it needs the cooperation of coarse, deep, semantic information and fine, shallow, local information [95]. FCN introduces skip layer (or skip connection) to accomplish that. Fig. 10 describes the skip layer used in FCN. In general, it up-samples feature maps from different deep layers with different scales, then add with feature maps in shallow layers, in this manner, the predictions can combine both global and local information.

2.3.2 U-Net

U-Net [120] is a popular segmentation network specially designed for medical imaging which is built upon FCN [95]. The detailed architecture of U-Net is shown in Fig. 11, it consists of a down-sampling (contracting) path and an up-sampling (expanding) path, this kind of architecture is also known as encoder-decoder. In the down-sampling path, image representations are extracted with successive convolution and pooling operations at different scales. After each down-sampling operation, the number of image features is doubled. In total, the down-sampling path has 5 convolution blocks with each has two 3×3 convolution layers with ReLU activation, followed by a max-pooling with

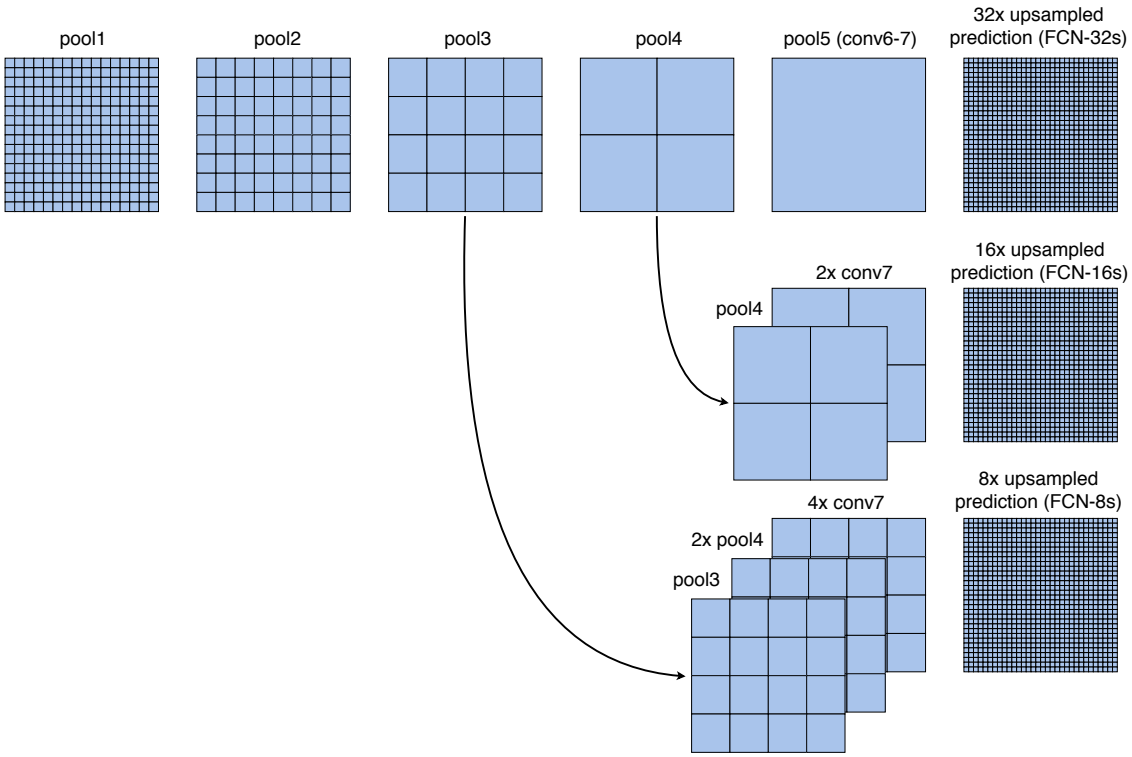

Figure 10: The skip connections in FCN. Pooling layers and prediction are shown as grids, convolution layers are omitted for clarity. Image is from [95].

stride 2×2 operation except the last block which is also called bottleneck block. While in the up-sampling path, the purpose of it is to recover resolution of the contextual information extracted from the down-sampling path and enable precise localization by utilizing the local information. Deconvolution operation with stride 2×2 is applied to up-sample the feature resolution, then concatenate with the features that have the same dimension from the down-sampling path, this is the skip connection in U-Net. After concatenation operation, two 3×3 convolution layers with ReLU activation are used to reduce the number of feature maps. Finally, a 1×1 convolution [93] is used to map the features to the desired number of segmentation classes.

In conclusion, U-Net has two major differences compared to FCN. Firstly, the architecture of U-Net is symmetric, it has a u-shaped structure. Secondly, U-Net applies concatenation operation instead of summation operation in FCN to fuse feature maps in skip connection. And the skip connection (or skip layer, residual connection [59]) in a CNN is extra connection between different layers that skips one or more layers.

The Overlap-tile Strategy

The resolution of medical image sometimes is extremely large. It's very challenging for training deep network with such large input images even with a modern GPU. U-Net [120] introduces a seamless patching strategy - the overlap-tile strategy. Basically, the whole image is divided into patches and

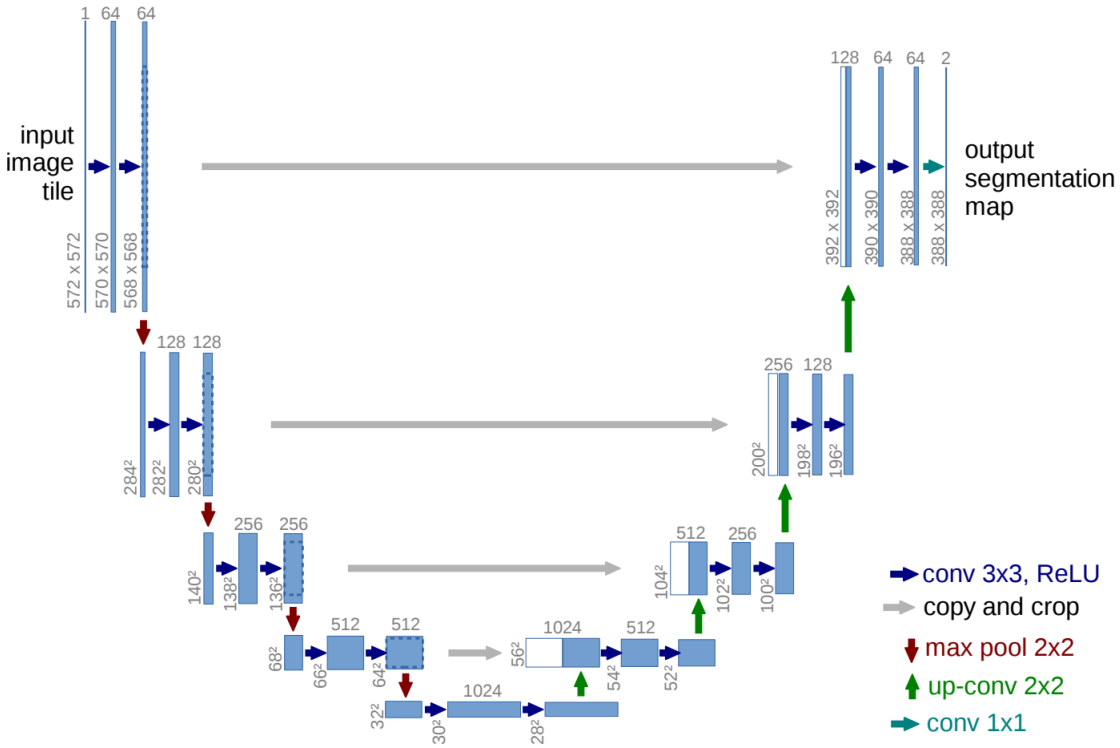

Figure 11: The U-Net architecture. Image is from [120].

all the patches are predicted one by one. In order to obtain prediction of a small part, we need image data from an area which is much more bigger than the small part of the input image. The explanation of this strategy is shown in Fig. 12. The area in green line of the input image is predicted using the area in blue line as the input. Image data is extrapolated by mirroring at image boundary.

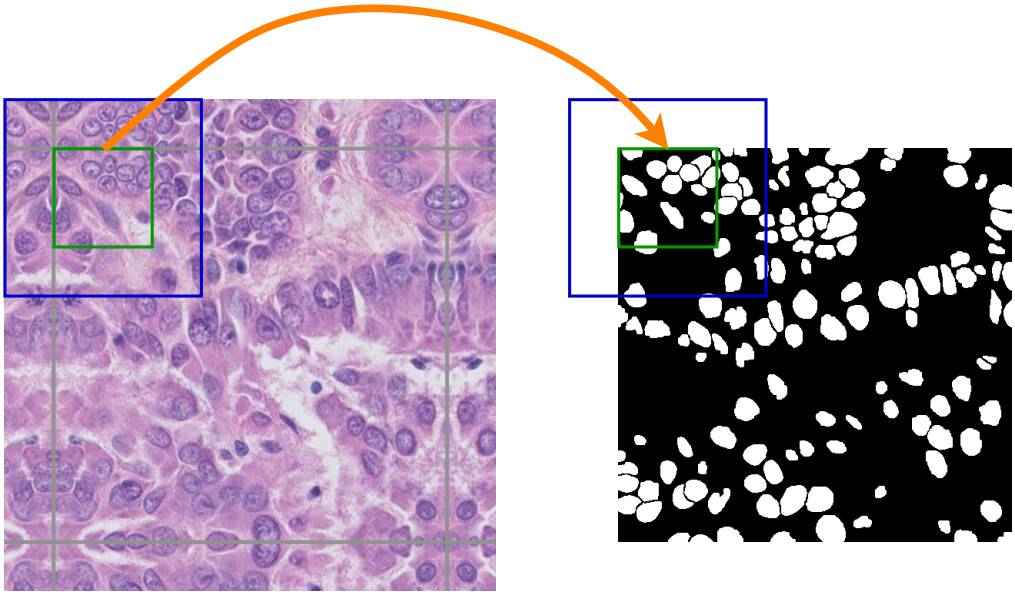

Figure 12: The Overlap-tile strategy. Left image is input image and right image is the corresponding segmentation mask. Images are from [107].

Chapter 3

Lung Segmentation in Chest X-ray by Fully Convolutional Networks

This chapter will present the first work of this thesis, the lung segmentation in chest X-ray by fully convolutional networks.

3.1 Introduction

A variety of imaging techniques are now available in the medical diagnosis field, such as X-ray imaging, computed tomography (CT) and magnetic resonance imaging (MRI). Despite the higher precise and sensitivity of CT and MRI, traditional X-ray imaging is still the most commonly used technique in medical diagnostic examinations and lung examinations because of less radiation dose and low cost. Chest radiography, as a cost-effective procedure and most widely used imaging techniques, is account for about one-third of all radiological procedures [141]. It provides a powerful tool to study various structures inside the thoracic cavity. Therefore chest radiography is widely used for the diagnosis of several diseases in clinical practices including emphysema, lung cancer and tuberculosis. Since the information extracted directly from lung regions such as size measurements, irregular shape and total lung volume can provide clues about early manifestations of life threatening diseases like emphysema [33, 103], pneumothorax, cardiomegaly and pneumoconiosis, accurate segmentation of lung regions in chest X-ray is a primary and fundamental step in computer-aided diagnosis (CAD) and plays a vital role for subsequent medical image analysis pipeline.

There are a number of difficulties and challenges for accurate lung segmentation in chest X-ray images. First of all, the shape and appearance of lung is greatly diverse due to differences in gender, age and health status. Secondly, the existence of external objects such as sternal wire, surgical clips and pacemaker will further makes the lung segmentation task much more difficult. Finally, some anatomical structures of lung may cause hardness for segmentation. For instance, the strong edges of the ribs and clavicle regions lead to local minima for many minimization methods.

3.2 Method

Most of the traditional segmentation methods for lung segmentation in chest X-ray rely on hand-crafted features. Recently, the progress of deep learning, especially CNNs based models have achieved huge success in many medical image analysis tasks.

In this study, we focus on applying robust deep CNN models to directly learn from image pixels for segmenting lung region in chest X-ray images. Specifically, we develop an automated framework based on FCN [95] and U-Net [120] for lung segmentation and demonstrates the superior performance of deep learning based approaches. Finally, we perform comparison study on 3 public chest X-ray image datasets to evaluate the performance of these models.

3.2.1 FCN

Since FCNs have 3 different architectures: FCN32, FCN16, FCN8, the only difference between these architectures is the skip connection. Specifically, FCN32 has no skip connection, FCN16 has one skip connection and FCN8 has two skip connections. In order to study the effect of skip connection, we adopt two FCNs for this study, FCN32 and FCN8. Following will give some details of these two architectures.

The architecture of FCN32 is shown in Fig. 13. It has five pooling layers, so the dimension of input image will be reduced to $\frac{1}{32}$ of the original input size, e.g. for an input image with size 512×512, the size in the smallest scale with be 16×16. Every level in FCN32 has two 3×3 convolution followed by ReLU activation except the last level. For the last level, it uses a 7×7 convolution with ReLU activation. After the 7×7 convolution, two 1×1 convolution with ReLU are used. Finally, it directly uses a transposed convolution with stride 32×32 to up-sample the 16×16 feature maps back to the original size, i.e. 512×512. Since the up-sampling rate is 32x, this type of FCN is called FCN32.

The architecture of FCN8 is shown in Fig. 14. Same with the FCN32, It also has five pooling layers. The major difference of FCN8 compared with FCN32 is that it uses feature addition operation to merge features in the previous layers. More specific, it firstly uses a transposed convolution with stride 2×2 to up-sample the 16×16 feature map back to 32×32, then a 1×1 convolution with ReLU is applied on the previous features map after the fourth pooling operation which has the same size 32×32, then uses addition operation to add these two feature maps with size 32×32. After addition operation, another transposed convolution with stride 2×2 is applied on the result feature maps. The feature map now has resolution 64×64, then add with another 64×64 feature map which is obtained from 1×1 convolution on the previous features after the third pooling operation. Finally, a transposed convolution with stride 8×8 is applied on the result feature after addition operation to obtain the final 512×512 segmentation map.

3.2.2 U-Net

The detailed network structure of U-Net is shown in Fig. 15. It is identical with the U-Net except for only one difference, in this study, we use convolution with padding instead of convolution without

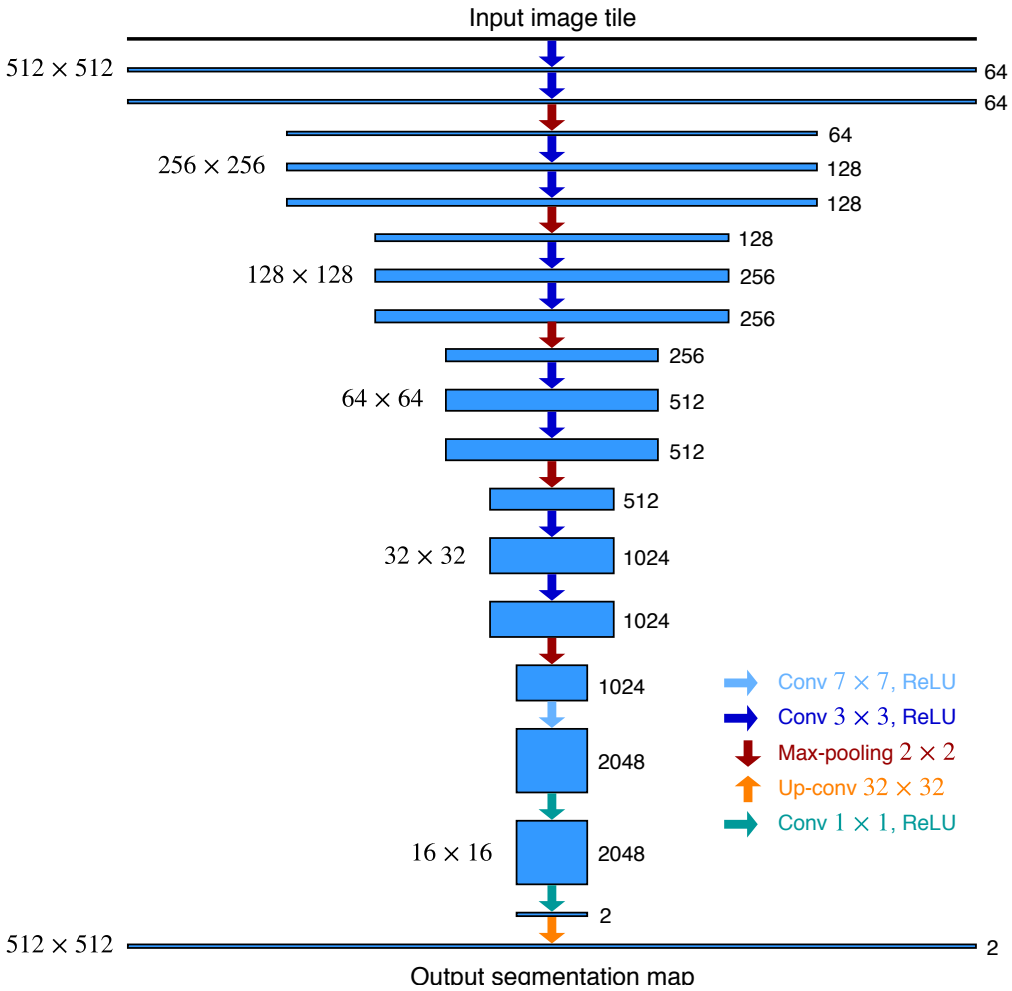

Figure 13: The architecture of FCN32 used in this study. Blue boxes represent image features. The number of features is indicated on the right of the box. The resolution of each level (features have the same resolution) is indicated on the left side of each level.

padding in the original U-Net. Therefore there is no dimension lose after every convolution operation.

3.3 Experimental Results

3.3.1 Datasets

Three publicly available datasets are used to evaluate the performance of different methodology in this study.

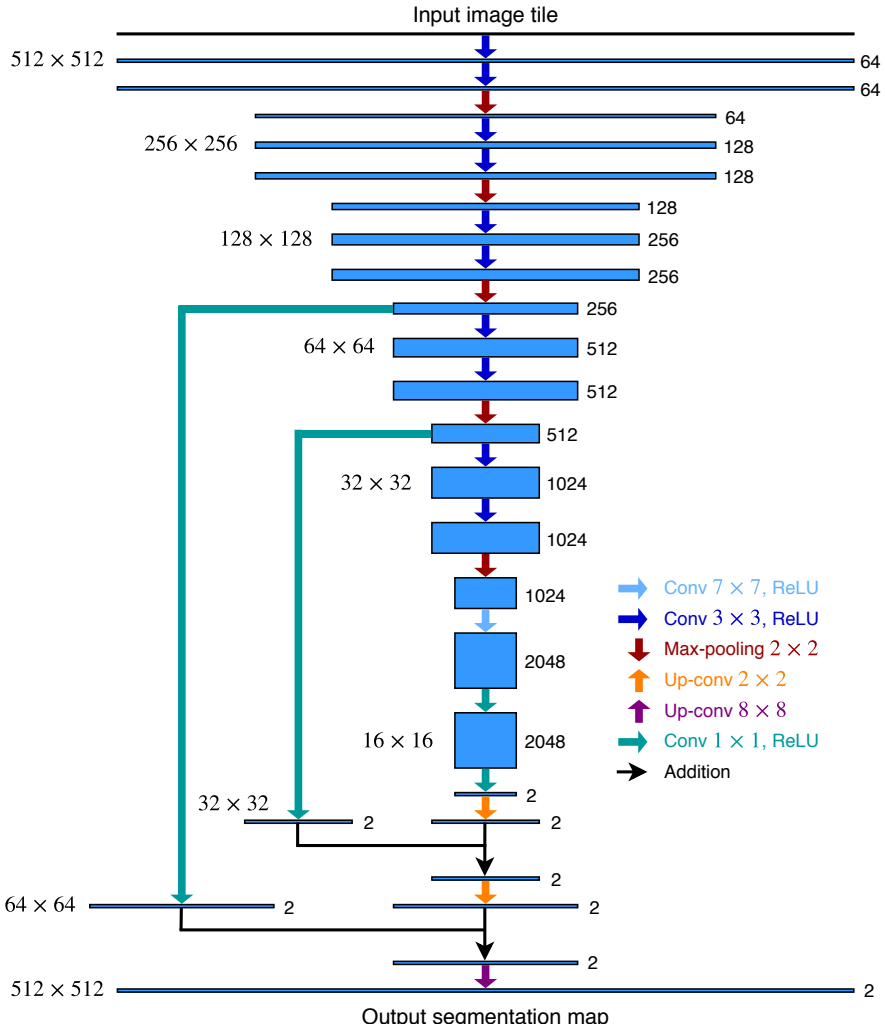

Figure 14: The architecture of FCN8 used in this study. Blue boxes represent image features. The number of features is indicated on the right of the box. The resolution of each level (features have the same resolution) is indicated on the left side of each level.

Montgomery County (MC) Dataset

The MC dataset [69] is from the department of Health and Human Services, Montgomery County, Maryland, USA. It contains 138 frontal chest X-ray images, among them 80 images are normal cases while 58 images are abnormal cases (i.e. tuberculosis). All images are provided in PNG format as 12-bit gray-scale images. The resolution of these images are either 4020×4892 or 4892×4020. The corresponding manual lung segmentation mask images are performed under the supervision of a radiologist.

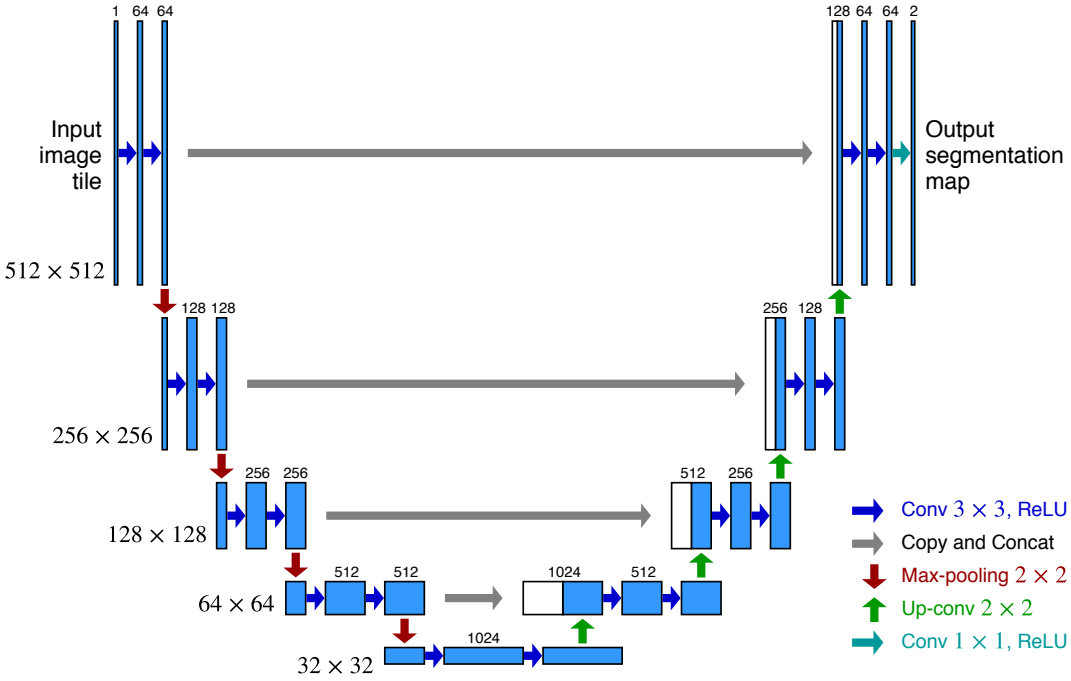

Figure 15: The architecture of U-Net used in this study. Blue boxes represent image features. The number of features is indicated on top of the box. The resolution of each level (features have the same resolution) is indicated at the bottom left of each level.

Shenzhen Dataset

The Shenzhen dataset [69] is from Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China. It consists of 662 frontal chest X-ray images in total, 326 images are normal cases while 336 images are abnormal cases. These images are also stored in PNG format and the resolution of them are vary but roughly 3000×3000. The corresponding lung segmentation masks are provided by [131]. However, for Shenzhen dataset, only 566 images have the corresponding manual lung segmentation mask images. Therefore only 566 images in this dataset are actually used for this study.

Japanese Society of Radiological Technology (JSRT) Dataset

The JSRT dataset [126, 142] is collected from 14 medical centers in Japan. It has 247 chest X-ray images, among them 93 images are normal cases and 154 are abnormal cases. All images are in PNG format and having 12-bit gray-scale with resolution 2048×2048. All the associated manual lung segmentation mask images are also available.

Three example chest X-ray images and their corresponding lung segmentation masks are shown in Fig. 16.

In addition, in order to evaluate the generalization ability of each segmentation model, we further merge all the 3 datasets which we call it Combined dataset in this study.

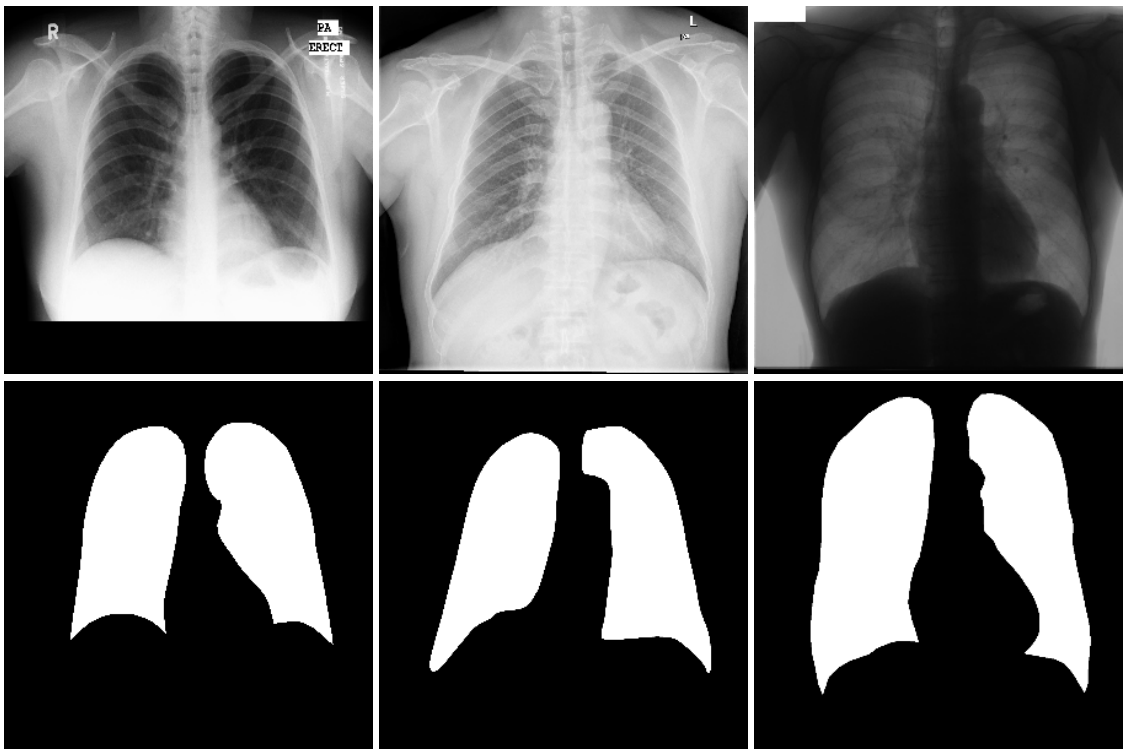

Figure 16: Example chest X-ray images and corresponding lung segmentation masks from 3 datasets (left: MC dataset, middle: Shenzhen dataset, right: JSRT dataset).

3.3.2 Evaluation Metrics

In order to evaluate the lung segmentation performance of different method and make a comparison in this study, we utilize 5 widely used evaluation criteria for medical image segmentation task, i.e. overlap measure (Overlap, also known as the Jaccard similarity coefficient), dice similarity coefficient (DSC), accuracy (ACC), specificity (SPE) and sensitivity (SEN, also known as recall).

Specifically, Overlap is the agreement between the segmented mask S and the ground truth segmentation mask GT over all pixels in the image, formally,

$$\text{Overlap} = \frac{|S \cap GT|}{|S \cup GT|} = \frac{|TP|}{|FP| + |TP| + |FN|} \tag{11}$$

where TP (True Positives) stands for pixels that are classified as foreground and are also foreground in ground truth. FP (False Positives) represents pixels that are classified as foreground but are background in ground truth. FN (False Negatives) means pixels that are classified as background but are foreground in ground truth.

DSC is the overlap between the segmented mask S and the ground truth segmentation mask GT, formally, it is defined as:

$$DSC = \frac{|S \cap GT|}{|S| + |GT|} = \frac{2|TP|}{2|TP| + |FN| + |FP|} \tag{12}$$

ACC is the proportion of the model's correct predictions. The definition of ACC is:

$$ACC = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{13}$$

where TN (True Negatives) represents pixels that are classified as background and are also background in ground truth.

Similarly, the SPE and SEN can be defined as:

$$\begin{aligned} SPE &= \frac{|TN|}{|TN| + |FP|} \\ SEN &= \frac{|TP|}{|TP| + |FN|} \end{aligned} \tag{14}$$

all the metrics mentioned above are the higher the better.

3.3.3 Implementation and Training Details

For each dataset, we use 70% and 10% data samples as the training and validation set, respectively. The remaining 20% data samples are used as the testing set. For the Combined dataset, the training/validation/testing splitting follows the same. The details of each dataset are shown in Table 1.

Table 1: Details of the chest X-ray image datasets used in this study.

Dataset	Training (70%)	Validation (10%)	Testing (20%)	Total (100%)
MC	97	14	27	138
Shenzhen	396	57	113	566
JSRT	173	25	49	247
Combined	666	95	190	951

Since the resolution of these images is diverse, we first re-size all the input chest X-ray images to 512×512 before passing them to the CNN model. In order to adjust image intensities for better image contrast, we also perform histogram equalization operation on the input chest X-ray images.

To conclude, we trained three models, i.e. FCN8, FCN32 and U-Net on these 4 datasets independently. In order to obtain stable performance results, for each model and each dataset, we perform 10 times running on 10 different random selected data split and average the final performance metrics results. All the models were trained by Adam optimizer with default suggested parameters [75]. The batch size for all the models is 4. We use the binary cross entropy loss as the loss function for all models. For the consideration of training efficiency and combat over-fitting, we use early stopping with patience 30 epochs, only the best model which has the lowest loss on validation set is used for evaluation on testing set.

Fig. 17, Fig. 18 and Fig. 19 show some example training curves including model accuracy and loss on these 4 datasets of FCN32, FCN8 and U-Net model, respectively. We can observe that all the three models are convergence after a few epochs.

Table 2: Lung segmentation results of different methods.

Method	Overlap (%) ↑	DSC (%) ↑	SEN (%) ↑	SPE (%) ↑	ACC (%) ↑
MC dataset					
Hybrid Nonrigid [17]	94.10	96.00	-	-	-
FCN32	90.36	94.77	94.82	98.48	97.57
FCN8	91.08	95.14	95.67	98.46	97.67
U-Net	**94.20**	**96.95**	**96.63**	**99.17**	**98.54**
Shenzhen dataset					
FCN32	90.14	94.74	94.33	98.47	97.42
FCN8	91.05	95.23	94.46	98.78	97.67
U-Net	**92.24**	**95.77**	**95.29**	**98.86**	**97.92**
JSRT dataset					
PC post [142]	94.50	-	-	-	-
ASM [125]	87.00	-	-	-	-
AAM [142]	84.70	-	-	-	-
ASM-OF [140]	92.70	-	-	-	-
Rule [2]	86.95	92.89	92.79	97.07	95.77
ShRAC [157]	90.70	-	-	-	-
SSAM [91]	93.09	96.41	95.25	98.88	97.69
FCN32	92.74	96.22	96.00	98.48	97.75
FCN8	94.16	96.98	96.98	98.71	98.20
U-Net	**94.97**	**97.46**	**97.07**	**99.09**	**98.49**
Combined dataset					
FCN32	90.65	95.04	95.52	98.10	97.43
FCN8	91.28	95.38	95.95	98.20	97.61
U-Net	**91.99**	**95.76**	**96.50**	**98.28**	**97.81**

3.3.4 Results and Discussions

Table 2 shows the segmentation performance results of different methods on MC, Shenzhen, JSRT and Combined dataset. We compare our CNN based approaches with some traditional segmentation techniques on the MC and the JSRT datasets, specifically, Hybrid Nonrigid [17] is a hybrid method based on nonrigid registration, PC post [142] is a pixel classification based method, ASM [125] is a method based on active shape model, AAM [142] is a method based on active appearance model, ASM-OF [140] is an active shape model with optimal features, Rule [2] is a rule-based segmentation approach, ShRAC [157] is an approach based on shape regularized active contour, and SSAM [91] is an approach based on statistical shape and appearance model. The performance results of these traditional approaches list in the Table 2 are directly from the respective literature. To the best of our knowledge, the Shenzhen dataset has not been studied in terms of segmentation research.

From these results, first of all, we can observe that all deep learning based methods achieve excellent performance in terms of all metrics on all the four datasets. This indicates that deep learning technique is particularly suitable for the lung segmentation task. Secondly, for all the four datasets, the performance of FCN8 is better than FCN32, this is the evidence that aggregating image features from different scales will make the segmentation results much better. Finally, from these obtained metrics, U-Net is much better than FCN8 and FCN32, the performance of U-Net

ranked the top among all models in all the datasets. It again highlights that for medical image segmentation, U-Net can achieve promising results.

Fig. 20, Fig. 21 and Fig. 22 show some example representative segmentation results comparison of these three models and the corresponding manual ground truth in MC, Shenzhen and JSRT dataset, respectively. From these visually comparison, we can observe that for some areas that are hardly segmented by FCN8 and FCN32, U-Net can segment these areas satisfactorily.

3.4 Conclusion

In this study, we focus on the task of lung segmentation in chest X-ray images. We apply FCN and U-Net on this task to demonstrate that the deep learning based approaches can achieve pretty good results. The experimental results on three public datasets and their combined dataset illustrate that U-Net achieved the best performance in terms of all metrics on all datasets for the lung segmentation task.

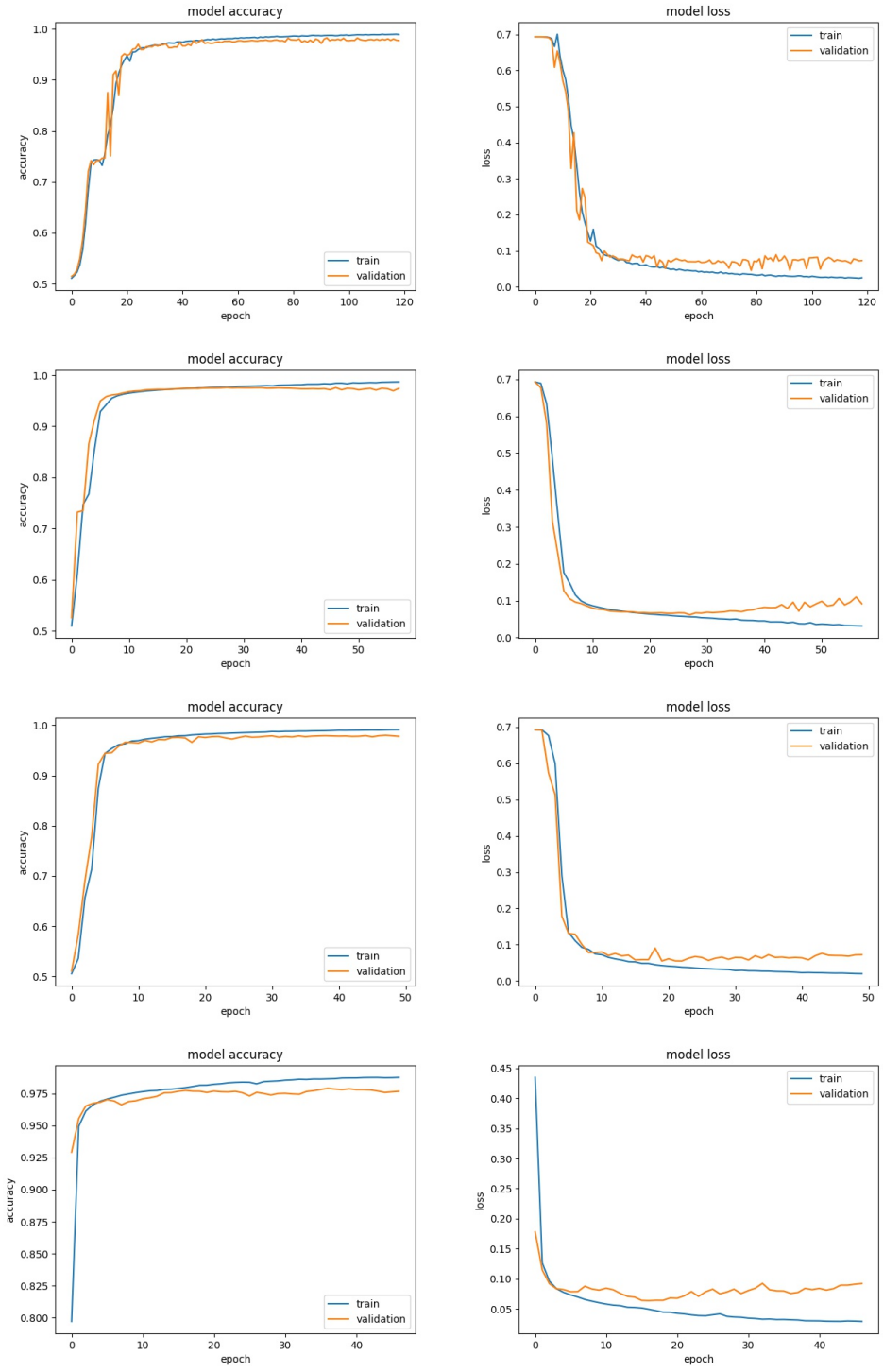

Figure 17: Training curves (accuracy and loss) of FCN32 on 4 datasets (From up to bottom: MC, Shenzhen, JSRT, Combined).

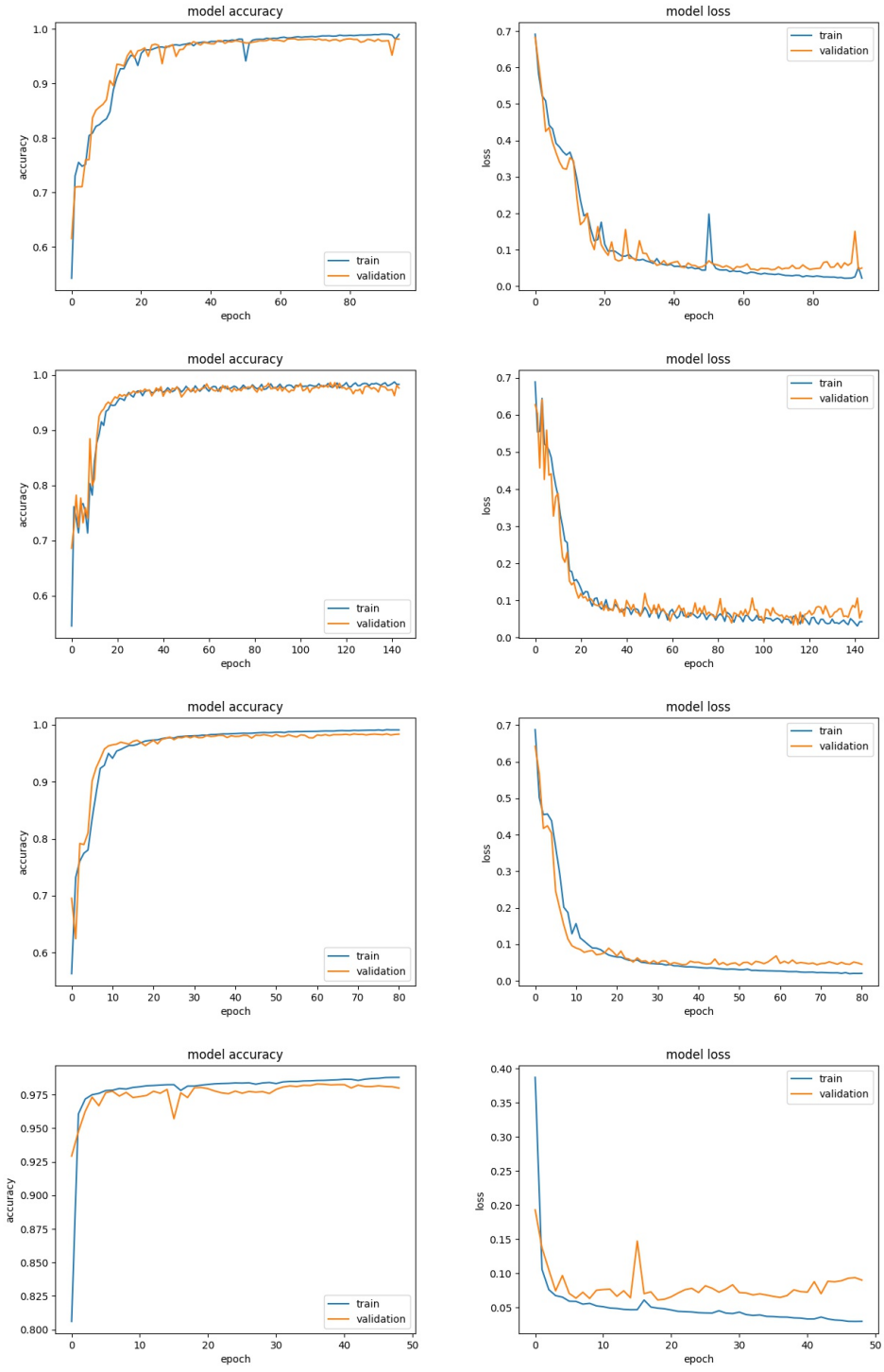

Figure 18: Training curves (accuracy and loss) of FCN8 on 4 datasets (From up to bottom: MC, Shenzhen, JSRT, Combined).

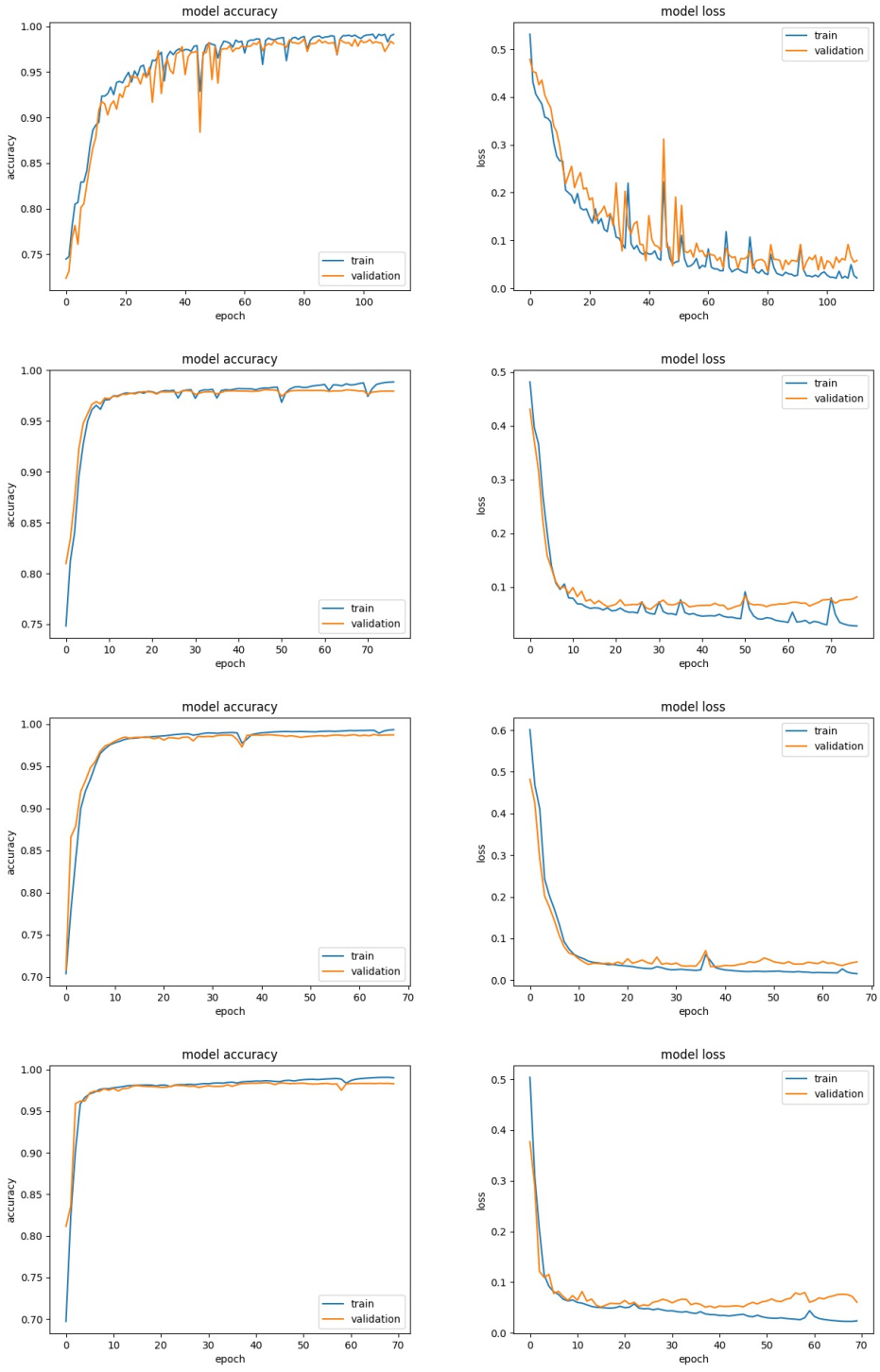

Figure 19: Training curves (accuracy and loss) of U-Net on 4 datasets (From up to bottom: MC, Shenzhen, JSRT, Combined).

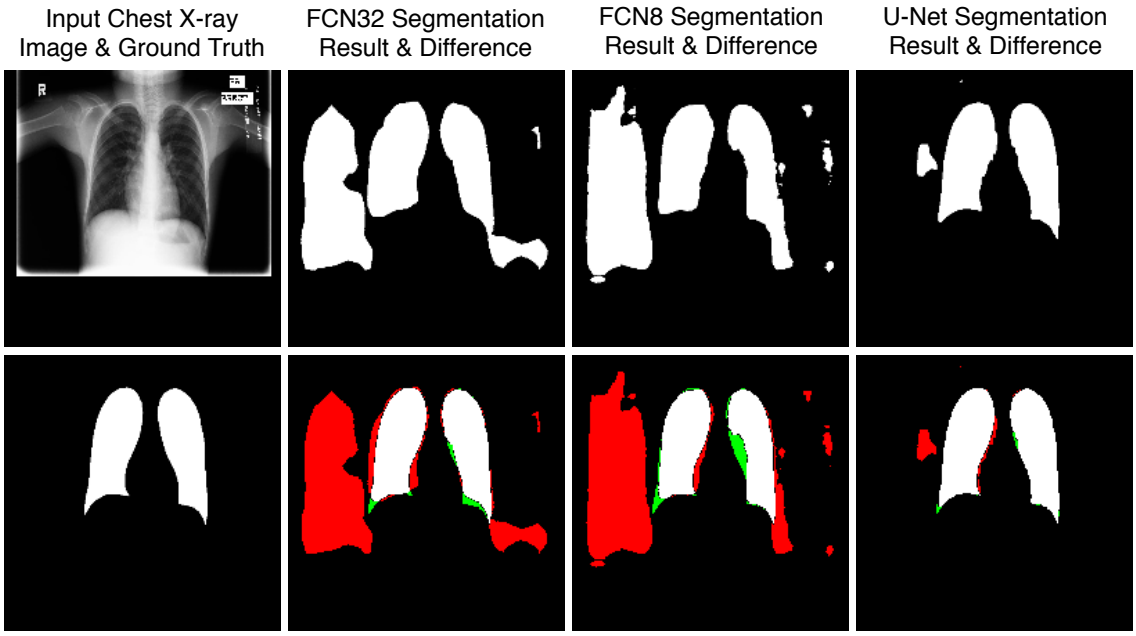

Figure 20: Sample segmentation results and their corresponding difference of different methods on the MC dataset. For the difference image, white color represents True Positives (TP), black color represents True Negatives (TN), red color represents False Positives (FP) and green color represents False Negatives (FN).

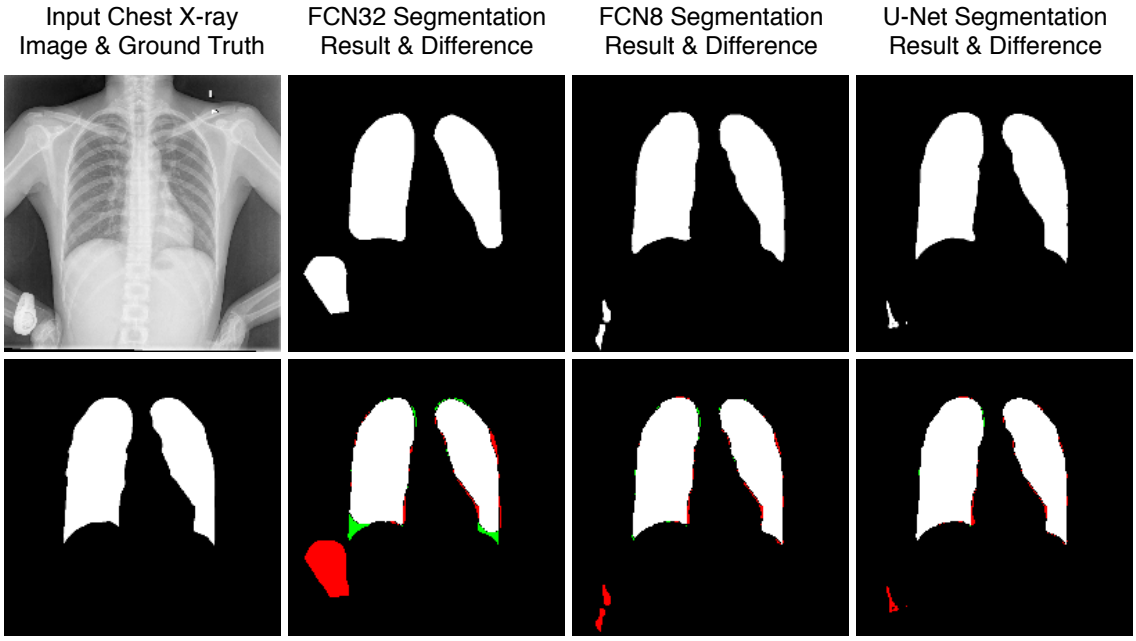

Figure 21: Sample segmentation results and their corresponding difference of different methods on the Shenzhen dataset. For the difference image, white color represents True Positives (TP), black color represents True Negatives (TN), red color represents False Positives (FP) and green color represents False Negatives (FN).

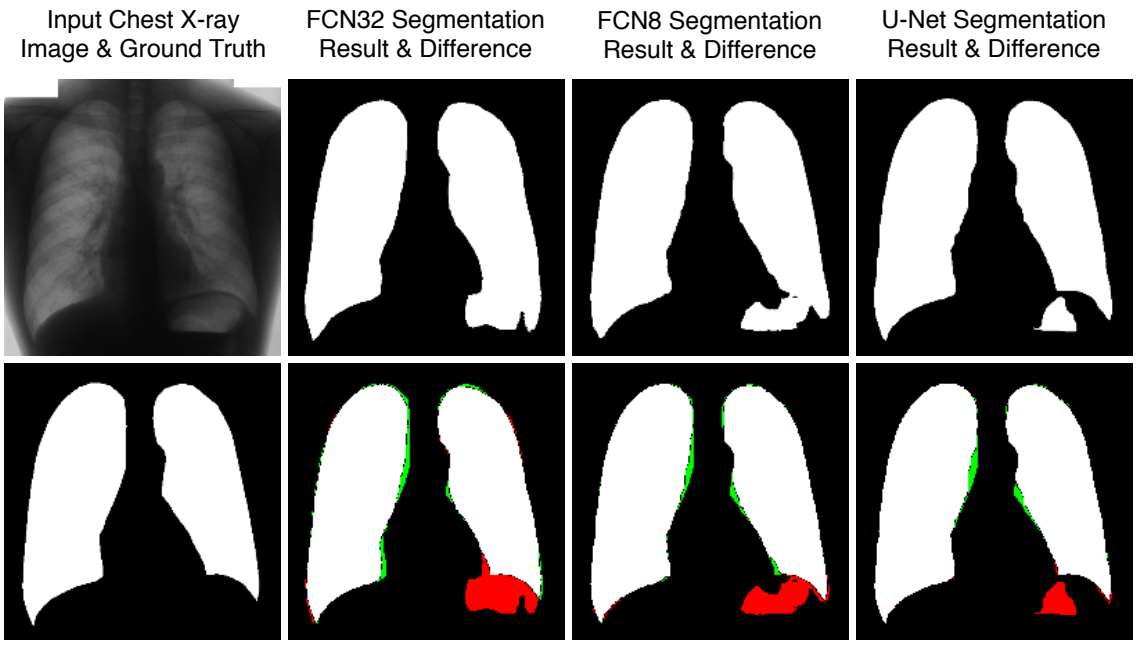

Figure 22: Sample segmentation results and their corresponding difference of different methods on the JSRT dataset. For the difference image, white color represents True Positives (TP), black color represents True Negatives (TN), red color represents False Positives (FP) and green color represents False Negatives (FN).

Chapter 4

A Two-Stage Learning Framework for Nuclei Segmentation

This chapter will present the second work of this thesis, a two-stage learning framework for nuclei segmentation in histopathological images.

4.1 Introduction

Histopathology plays a critical role in the understanding, prognosis, diagnosis and treatment of almost all discovered diseases [127]. The histopathological image data of a patient can be checked by a Pathologist in order to determine following treatment. Tissue slides are informative for many diseases such as cancer grade and sub-type. The studies of nuclear distribution, morphometric and appearance features in tissue slides provide important clues in clinical practice. Since histopathological images provide extensive information regarding cell morphology and tissue architecture, they are used in a broad range of applications in clinical practice, e.g., medical diagnosis [53], cancer malignancy grading [26] and treatment effectiveness prediction [43]. Moreover, nucleus contains a large number of epigenetic and genetic codes that can control and regulate cell type, morphology and function. With the advent of cellular staining methods such as hematoxylin and eosin (H&E) in which some useful specific structures such as cells, cell nuclei and collagen are highlighted, the interpretation and determination of abnormal phenotypes in these stained tissue slides has been interpreted by human that is prone to be subjective and time-consuming.

Digital pathology has been widely studied from both image analysis researchers and pathologists, especially with the introduction of whole-slide imaging (WSI). WSI is a technology that can digitally capture images that represent the whole stained tissue from a glass slide in a high-speed and high-resolution way [41]. The advantage of WSI is not only provides a convenient way to store and share these digitized tissue slides, but also paves a way for analyzing these informative images automatically using image analysis techniques. Specific to histopathological images, digital histopathological image analysis aims to automatically analyze histopathological images, which can significantly improve the

reproducibility and objectivity of diagnosis [53]. Segmentation of nuclei in stained tissue images is a fundamental and essential step for interpreting and analyzing these images. Accurate and robust nuclei segmentation is a key pre-requisite step for Computer Aided Diagnosis (CAD). The purpose of nuclei segmentation is obtain the detailed contours of each nucleus and separate individual nuclei from input images for further processing.

4.2 Motivation

Although much progress have been made for nuclei segmentation in histopathological images, there are still several challenges associated with this task. First of all, nuclear appearance, morphometric, shape, color and density may varies by different organs. Fig. 23 gives an example of this variation from 5 different organs: liver, colon, prostate, stomach and kidney. Secondly, some nuclei which has extremely small size may very difficult to detect and segment, for instance, the nuclei of kidney in Fig. 23. Finally, as shown in Fig. 24, touching and overlapping nuclei are also especially difficult to segment.

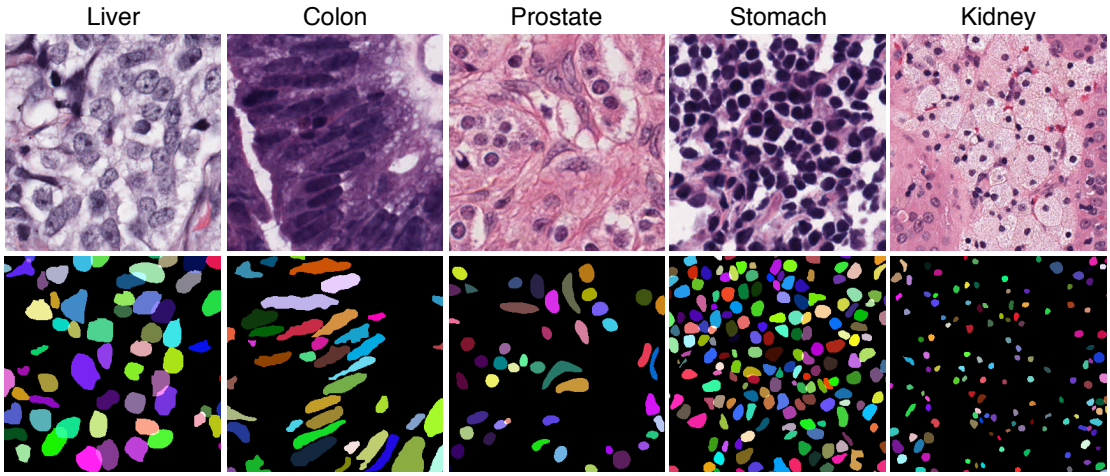

Figure 23: Examples of H&E stained images (up) and corresponding nuclei segmentation map (bottom) for different organs (columns). Images are from [81].

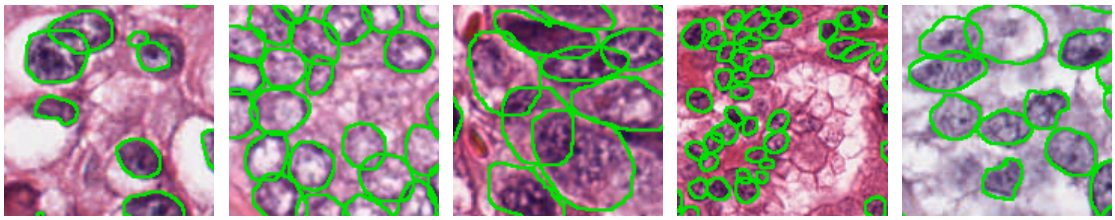

Figure 24: Examples of overlapping and touching nuclei, green lines outline the boundary of each nuclei in H&E stained images. Images are from [81].

In conclusion, nuclei segmentation still has several challenges such as difficulty in segmenting

the overlapping and touching nuclei, detecting the nuclei which have small size, and limited generalization ability to different organs and tissue types. In order to tackle these challenges, many approaches have introduced the nuclei-boundary prediction as part of nuclei segmentation to help segment overlapping and touching nuclei. Kumar *et al.* [81] propose a CNN model that predicts nuclei and boundary segmentation map based on a patch-wise approach. Naylor *et al.* [108] convert the binary segmentation task into a regression task by predicting the distance map of nuclei. Although these methods have lead to some performance improvements, they still have some disadvantages such as needing complex post-processing and excessive redundant computation.

4.3 Background

4.3.1 Stacking U-Nets

There are many works attempt to make some improvements based on FCN or U-Net [38, 102, 90, 27, 11, 121, 46, 74]. Drozdzal *et al.* [38] extend FCN by adding short skip connections from residual networks [58, 59] and demonstrates that these short connections together with the long skip connections originally in FCN and U-Net can alleviate the gradient exploding/vanishing problem and enable us to build very deep networks for image segmentation. Milletari *et al.* [102] propose a 3D FCN for dealing with 3D volumetric image segmentation in medical imaging. Li *et al.* [90] utilize the dense connections in DenseNet [63] to design a 2D DenseUNet and a 3D counterpart to propose a novel hybrid densely connected U-Net. On the other hand, the idea of cascading or stacking multiple FCNs or U-Nets has attracted intensive and exhaustive research. Christ *et al.* [27] cascade two FCNs and design a dense 3D conditional random fields for liver and lesion segmentation in CT abdomen images. Sevastopolsky *et al.* [121] stack two kinds of building blocks, U-Net or Res-U-Net which is U-Net extends with residual connection, for optic disc and cup image segmentation. Ghosh *et al.* [46] stack multiple U-Nets extend with dilated convolution [155] for ground material segmentation in remote sensing images. Khalel *et al.* [74] use a stack of U-Nets for object segmentation in aerial imagery. However, these stacking approaches mentioned above do not design different tasks or outputs for the sub-networks, i.e., in each case, their stacked networks perform exactly the same task with the same output. The segmentation network designed by Bi *et al.* [11] has multiple outputs but again performs the same task.

4.3.2 Deep Layer Aggregation

U-Net originally has skip connections to fuse image representations or features in different levels. However, only features in the up-sampling path are merged with the corresponding features from the down-sampling path, even in U-Nets extend with residual connections [38, 121] or dense connections [90], they only merge features stay in the same level or have the same resolution. Thus may not integrate the extracted features of different levels in an effective manner. In summary, the original skip connections in U-Net are still linear and shallow [156]. Yu *et al.* [156] introduce two kinds of Deep Layer Aggregation (DLA): iterative deep aggregation (IDA) and hierarchical deep aggregation

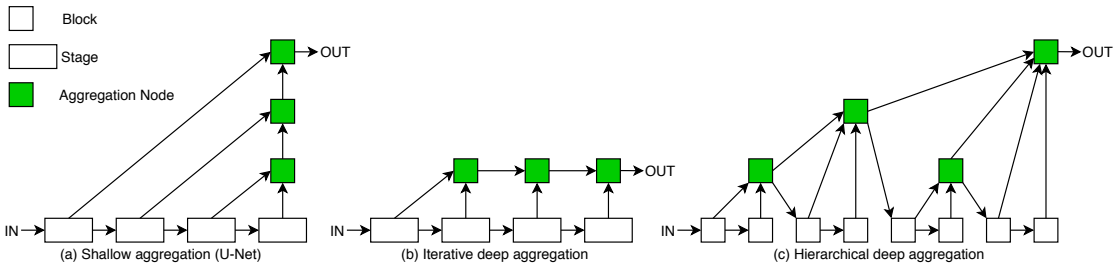

Figure 25: The shallow aggregation, IDA and HDA. Image is from [156].

(HDA) to better fuse image representations across different levels. Specifically, connections are extended in iterative (IDA) and hierarchical (HDA) manner to exploit the global (coarse) and local (fine-grained) information. Fig. 25(a), Fig. 25(b) and Fig. 25(c) show the structure of shallow aggregation which was used in U-Net, the structure of IDA and HDA, respectively.

4.3.3 Curriculum Learning

Curriculum learning was proposed by Bengio *et al.* [8]. The main motivation of curriculum learning is to imitate the characteristics of human learning, from simple to difficult to learn the curriculum (in the machine learning context, it can point to easy samples and hard samples), so that the model can easily find better local optimum, and accelerate the speed of training. Curriculum learning can be interpreted as a continuation method, it starts with easier or simpler concepts and progress to more complex or hard concepts that depend on the previous learned easier concepts [48]. Curriculum learning has been successful in numerous computer vision tasks [80, 88, 132].

4.4 Methodology

In this section, we give more details of our proposed approach for nuclei segmentation.

4.4.1 Overview

Inspired by the core idea of curriculum learning [8] and the aforementioned segmentation approaches, we propose a novel nuclei segmentation approach based on a two-stage learning framework to solve the above-mentioned challenges in nuclei segmentation. The core idea of curriculum learning is that a complex task can be solved by dividing it into numerous sub-tasks, and one can start with the easiest one, followed by subsequent tasks that have increased level of difficulty. Specifically, in order to tackle small, overlapping and touching nuclei, we convert the original binary segmentation task into a two-step task by adding the prediction of nuclei-boundary (3-classes) as an intermediate step. Along with this two-step task, we design a two-stage learning framework by stacking two U-Nets that have two different outputs. The coarse boundary from the first stage acts like auxiliary information to guide the segmentation of small, overlapping and touching nuclei in the second stage, therefore decreases the difficulty of segmenting nuclei directly from input images. In addition, we extend

our U-Net with DLA, which has been demonstrated to have superior performance in many visual applications.

Our segmentation network is an end-to-end learning framework. The only pre-processing step is just color normalization and no post-processing step is needed. After color normalization, image patches are seamlessly extracted by the overlap-tile strategy and then fed into our network. During prediction phase, the predicted image patches of our network are merged together to obtain the final segmentation map thus make our model can handle arbitrary size of images.

4.4.2 Color Normalization

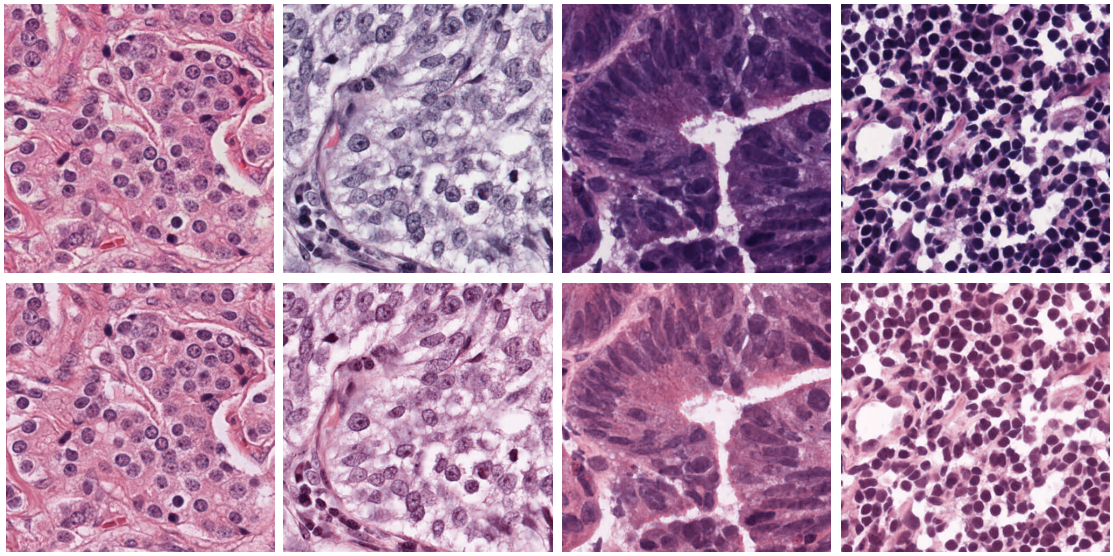

Figure 26: Example image samples (up) and their corresponding color-normalized image samples (bottom). The first column is the target image. Images are from [81].

H&E stain is one of the most widely used stains in histopathological images and is usually the gold standard for medical diagnosis. However, H&E stained histopathology images in general have diverse color variations due to differences in scanners, materials and staining process, therefore color normalization techniques are widely used to eliminate color variations and preserve tissue structures. Vahadane *et al.* [139] proposed a color normalization method based on sparse non-negative matrix factorization (SNMF) and achieved superior performance. We adopt the technique for color normalization and the target image was chosen by the recommendation of the dataset [81]. Fig. 26 shows some examples of color normalization.

4.4.3 Network Architecture

Fig. 27 illustrates the detailed architecture of our proposed network for nuclei segmentation. In the first stage, a U-Net with DLA is utilized to predict the 3-classes nuclei-boundary segmentation map from the color-normalized image patches. The first stage consists of a down-sampling path and an up-sampling path, just the same as the U-Net. For the second stage, a much more light-weighted and

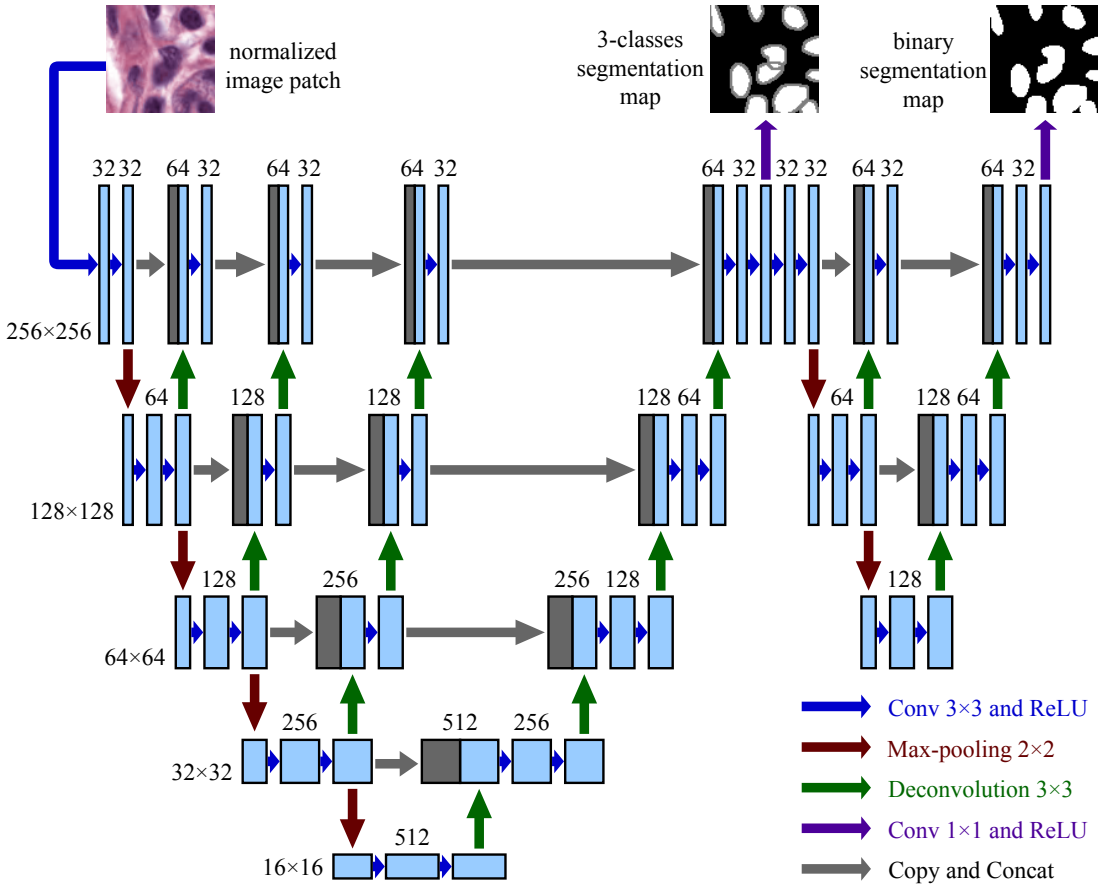

Figure 27: The architecture of our proposed segmentation network. Blue boxes represent image features. The number of features is indicated on top of the box. The resolution of each level (features have the same resolution) is indicated at the bottom left of each level.

shallow U-Net with DLA is used to refine the coarse nuclei-boundary segmentation map generated from the first stage for the final binary segmentation map. By a light-weighted and shallow U-Net, we mean it only has two max-pooling layers compared to four in the original U-Net. Based on our experiments, we did not notice considerable performance difference between deep and shallow architectures for the second stage, so for the consideration of computational cost and efficiency, the shallower one is used. The input of the second stage is the feature maps in the first stage before 1×1 convolution and the output is the binary segmentation map.

Formally, let I be the input color-normalized image patch, therefore I belongs to RGB image domain, $I \in \Omega = \mathbf{R}^{w \times h \times 3}$, where w and h indicate image width and height, respectively. Let S_1 be the nuclei-boundary segmentation map achieved from the first stage $S_1 \in \Psi = \{0, 1, 2\}^{w \times h}$, and S_2 be the binary segmentation map obtained from the second stage $S_2 \in \Phi = \{0, 1\}^{w \times h}$. The task of the first stage t_1 can be defined as

$$t_1 = \Omega \rightarrow \Psi$$

and the task of the second stage t_2 can be defined as

$$t_2 = \Psi \rightarrow \Phi$$

These two tasks are trained simultaneously in an unified network model.

Following [156], IDA is defined as

$$I(x_1, ..., x_n) = \begin{cases} x_1 & \text{if } n = 1 \\ I(N(x_1, x_2), ..., x_n) & \text{otherwise} \end{cases} \tag{15}$$

where the aggregation node is denoted as N. In our case, N is defined as

$$N(x_1, x_2) = \text{Conv}(\text{Concat}(x_1, x_2)) \tag{16}$$

where Conv is a 3×3 convolution operation followed by ReLU activation, and Concat represents the concatenation operation.

As illustrated in Fig. 27, deconvolution operation is applied to every last image features at each level in the down-sampling (encoder) path, then merged iteratively with the features at previous levels. After that, 3×3 convolution and ReLU is used in order to keep the same feature number at each scale. Finally, these feature maps in each different scale after IDA still merge with the corresponding feature maps in up-sampling (decoder) path. Same with the original U-Net, we use 1×1 convolution to reduce the number of features and softmax activation function to generate the segmentation map.

4.4.4 Loss Function

Since our segmentation network has two outputs - one is the 3-classes nuclei-boundary segmentation map for the first stage and another one is the final binary segmentation map for the second stage, we have two loss functions. The loss function of each stage is the categorical cross entropy loss:

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{C} I_{i,k} \log p_{i,k} \tag{17}$$

where N is the total number of image pixels and C is the number of segmentation categories. The term $I_{i,k}$ is the indicator function of whether the i-th pixel belongs to the k-th category. The $p_{i,k}$ is the probability predicted by the model for the i-th pixel belonging to the k-th category. The overall loss $\mathcal{L}$ of the network is the weighted summation of these two loss terms of the two stages,

$$\mathcal{L} = \alpha \mathcal{L}_1 + (1 - \alpha) \mathcal{L}_2 \tag{18}$$

where α is the weight such that $0 \leq \alpha < 1$, and $\mathcal{L}_1$ and $\mathcal{L}_2$ are the losses of the first and second stage, respectively. The weight α is a hyper-parameter and will be tuned in experiments.

4.5 Experiments and Results

This section will give the details of experiments and the performance results our proposed approach.

4.5.1 Datasets

In order to make a comparison between other nuclei segmentation methods, we evaluate our proposed approach on two publicly available nuclei datasets.

The first dataset is proposed in [81], it contains 30 H&E stained histopathology images and each image has 1000×1000 resolution. These images are captured at 40x magnification from The Cancer Genomic Atlas (TCGA) archive and taken from seven different organs (breast, liver, kidney, prostate, bladder, colon and stomach). In total, more than 21000 nuclei are annotated in this dataset. According to the training and testing protocol suggested by [81], 30 images are split into three subsets, 12 images for training, 4 images for validation and 14 images for testing. In addition, in order to test the generalization ability to images taken from organs that do not appear in the training set, it further divides the testing set into two testing sets: same organ testing and different organ testing. The images in the different organ testing set are taken from organs not represented in the training set - bladder, stomach and colon. The details of this dataset are shown in Table 3. This dataset will be referred as TCGA for convenience.

Table 3: Composition of the TCGA dataset and the associated training/testing protocol.

Data subset		Nuclei	Images							
		Total	Total	Breast	Liver	Kidney	Prostate	Bladder	Colon	Stomach
Training		9669	12	3	3	3	3	-	-	-
Validation		3703	4	1	1	1	1	-	-	-
Testing	Same organ testing	4130	8	2	2	2	2	-	-	-
	Different organ testing	4121	6	-	-	-	-	2	2	2
Total		21623	30	6	6	6	6	2	2	2

The second dataset is proposed in [14, 15]. It consists of 50 H&E stained tissue images with 512×512 resolution and totally 4022 nuclei have been annotated. The maximum of number of nuclei in one image is 293 and the minimum number is 5, with an average of 80 nuclei per image and standard deviation of 58. All the images are taken from 11 Triple Negative Breast Cancer (TNBC) patients, and include different cell types such as myoepithelial breast cells, endothelial cells and inflammatory cells. This dataset will be referred as TNBC for convenience. Fig. 28 shows some example images and annotations from the TNBC dataset.

4.5.2 Evaluation Metrics

Two types of metrics are used to evaluate the performance of different approaches in this study: object-level and pixel-level metrics.

The Aggregated Jaccard Index (AJI) presented in [81] is used as an object-level evaluation metric. Basically, the AJI is an extension of the Jaccard Index. Specifically, the AJI is defined as

$$AJI = \frac{\sum_{i=1}^{K} |GT_i \cap PD_j^*(i)|}{\sum_{i=1}^{K} |GT_i \cup PD_j^*(i)| + \sum_{l \in U} |PD_l|} \tag{19}$$

where $GT = \cup_{i=1,2,\cdots,K} GT_i$ are the pixels of whole ground truth nuclei objects, and $PD = \cup_{j=1,2,\cdots,L} PD_j$ are the pixels of whole predicted nuclei objects. $PD_j^*(i)$ is the connected component object from the predicted result that has the maximum Jaccard Index with the ground truth

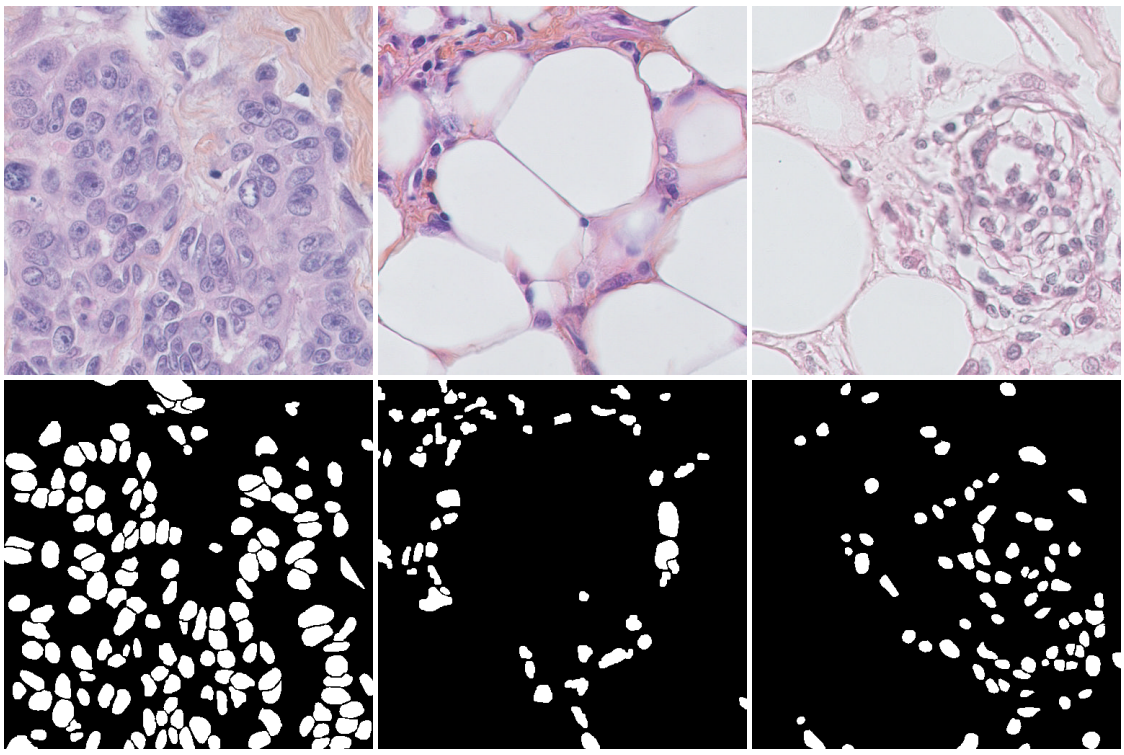

Figure 28: Example images (up) and associated ground truth segmentation masks (bottom) of the TNBC dataset.

objects, and U is the union of predicted nuclei objects that does not correspond to any ground truth objects (also known as ghost objects).

For pixel-level evaluation metrics, we use precision, recall and F_1 score. These 3 metrics are defined as

$$\begin{aligned} precision &= \frac{TP}{TP+FP} \\ recall &= \frac{TP}{FN+TP} \\ F_1 &= 2 \cdot (\frac{precision \cdot recall}{precision + recall}) \end{aligned} \tag{20}$$

where TP is true positives, FP is false positives and FN is false negatives.

4.5.3 Implementation Details

Considering the resolution of some images is 1000×1000 and it's very hard to process with a GPU, we divide the input H&E images into small patches. For the consideration of performance and GPU memory limitation, the size of image patch in our experiments is 256×256. The nuclei-boundary mask images for training are obtained by image morphological operations on the ground truth of segmentation maps. Specifically, the difference image of the dilation result of a mask image and the erosion result of a mask image can be used as the nuclei-boundary mask. Fig. 29 gives some examples of the nuclei-boundary masks.

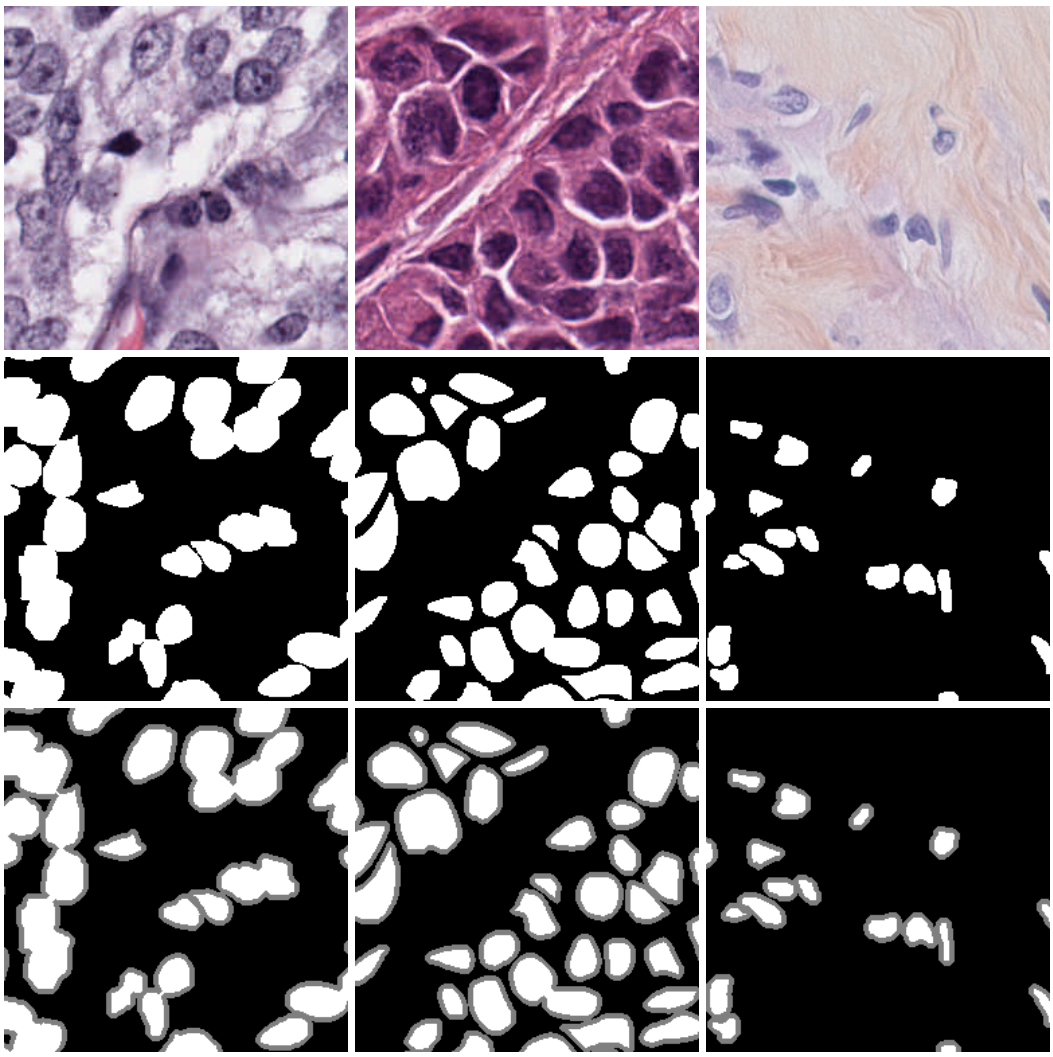

Figure 29: Example images (up) and associated ground truth binary masks (middle) and associated ground truth nuclei-boundary masks (bottom). The first two columns are from the TCGA dataset while the last one from the TNBC dataset.

Since we propose two different types of improvements in our approach compared to other segmentation methods, in order to validate each of them, we implement three models in this study, i.e., **U-Net (DLA)** which is a U-Net extended with DLA, **Ours** which is our two-stage learning model with the original U-Net and **Ours (DLA)** which combines two-stage learning and DLA. All the three models are trained with the same configuration. We use an ADADELTA [158] optimizer with the default values suggested by Zeiler [158] to train all the three models, all the weights were initialized with MSRA initialization [57] and trained from scratch. For training efficiency and combat over-fitting, we use early stopping with patience 30, only the best model which has the lowest loss on validation set is used for evaluation on testing set.

4.5.4 Results and Discussions

First of all, in order to determine the loss weight α in equation 18, we perform experiments on the TCGA dataset using **Ours (DLA)** model and the results are shown in Table 4. From Table 4, we can observe that when α is set to 0.8, both F_1 score and AJI achieve the highest value. Therefore, for the rest of our experiments, the loss weight α is set to 0.8.

Table 4: Results by choosing different loss weight α.

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F_1 Score	0.793	0.797	0.800	0.800	0.801	0.804	0.804	0.805	**0.808**	0.806
AJI	0.567	0.571	0.571	0.575	0.578	0.581	0.581	0.586	**0.590**	0.586

For the TCGA dataset, in order to make a comparison with other methods, we follow the same training and testing split suggested by Kumar *et al.* [81]. We compare our methods with numerous standard segmentation architectures such as FCN-8 [95], Mask R-CNN [56], U-Net [120] and other state-of-the-art nuclei segmentation methods like DIST [108] and CNN3 [81]. In addition, we also stack two U-Nets, where both sub-nets have the same binary segmentation task, which we call Stacked U-Net. In Table 5 and Fig. 30 we show the AJI for each organs of different methods on the TCGA test set. While Table 6 and Fig. 31 show F_1 scores for different organs of different methods on the TCGA test set. From the results on the TCGA dataset, we can observe that our model rank the top for overall in terms of both AJI and F_1 score. For overall, the highest AJI of our model is nearly 3 percent higher than the highest AJI achieved by other methods. The highest F_1 score of our model is roughly 1.5 percent higher than the highest F_1 score obtained by other methods. Specific to each organ, our models achieve the highest AJI in 5 out of 7 organs: bladder, colorectal, stomach, breast and kidney, second in liver, third in prostate. On the other hand, in terms of F_1 score, our models rank the top in 4 out of 7 organs: bladder, colorectal, breast and kidney, second in stomach, liver and prostate.

Table 5: AJI of different methods on the TCGA test set.

Aggregated Jaccard Index (AJI) ↑								
Organ	Bladder	Colorectal	Stomach	Breast	Kidney	Liver	Prostate	Overall
FCN-8 [95]	0.5376	0.4018	0.5279	0.5509	0.5267	0.5045	0.5709	0.5171
Mask R-CNN [56]	0.5011	0.3814	0.6151	0.4913	0.5182	0.4622	0.5322	0.5002
U-Net [120]	0.5403	0.4061	0.6529	0.4681	0.5426	0.4284	0.5888	0.5182
CNN3 [81]	0.5217	0.5292	0.4458	0.5385	0.5732	**0.5162**	0.4338	0.5083
DIST [108]	0.5971	0.4362	0.6479	0.5609	0.5534	0.4949	**0.6284**	0.5598
Stacked U-Net	0.6138	0.5188	0.5845	0.5605	0.5647	0.4594	0.5300	0.5474
U-Net (DLA)	0.6215	0.5322	0.5938	0.5747	0.5624	0.4642	0.5602	0.5584
Ours	0.6263	0.5346	0.6352	0.6037	0.5928	0.4961	0.5606	0.5784
Ours (DLA)	**0.6285**	**0.5376**	**0.6620**	**0.6096**	**0.6024**	0.5142	0.5720	**0.5895**

In addition, we also perform the same/different organ testing protocol described in [81], the AJI and F_1 scores of different methods are shown in Table 7. From the table, we can see our model obtained the highest AJI and F_1 score both on the same organ testing set and the different organ testing set. One interesting thing is that our models perform better on different organ testing set

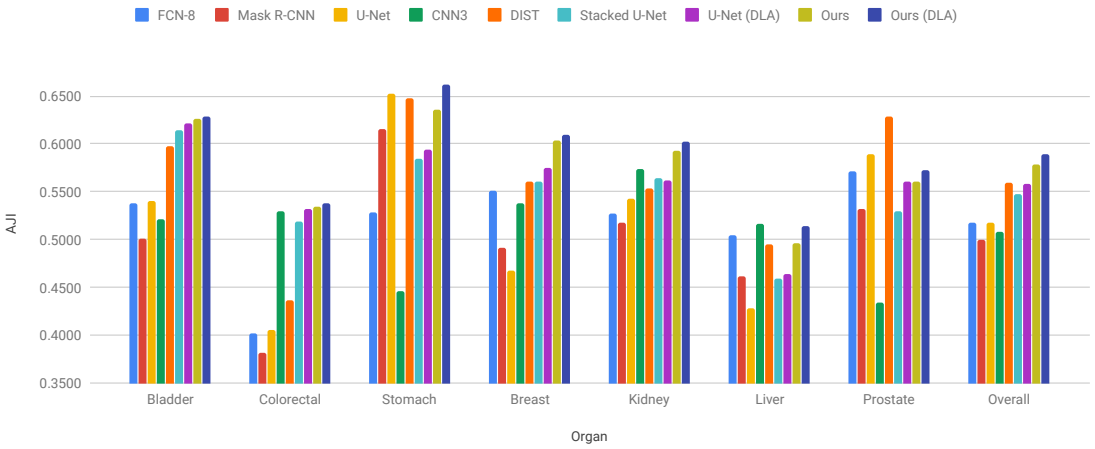

Figure 30: Comparative analysis of AJI for each organs on the TCGA test set.

Table 6: F_1 scores of different methods on the TCGA test set.

F_1 Score ↑								
Organ	Bladder	Colorectal	Stomach	Breast	Kidney	Liver	Prostate	Overall
FCN-8 [95]	0.8084	0.6934	0.7982	0.8113	0.7597	**0.7589**	**0.8367**	0.7809
Mask R-CNN [56]	0.7610	0.6820	0.8269	0.7481	0.7554	0.7157	0.7401	0.7470
U-Net [120]	0.7953	0.7360	0.8638	0.7818	0.7913	0.6981	0.7904	0.7795
CNN3 [81]	0.7808	0.7399	**0.8948**	0.7181	0.7222	0.6881	0.7922	0.7623
DIST [108]	0.8196	0.7286	0.8534	0.8071	0.7706	0.7281	0.7967	0.7863
Stacked U-Net	0.8249	0.7685	0.8498	0.7990	0.7986	0.7276	0.7829	0.7930
U-Net (DLA)	0.8296	0.7756	0.8530	0.8025	0.7994	0.7296	0.7895	0.7970
Ours	0.8213	0.7773	0.8700	0.8068	**0.8066**	0.7437	0.7890	0.8021
Ours (DLA)	**0.8360**	**0.7808**	0.8629	**0.8183**	0.8022	0.7513	0.8037	**0.8079**

compared to same organ testing set, one of the reasons may be that segmenting the nuclei in the different organ testing set is much easier than segmenting the nuclei in the same organ testing set, this phenomenon happened on other methods too (Stacked U-Net, DIST and U-Net).

For the TNBC dataset, we follow the same leave-one-patient-out scheme used by Naylor *et al.* [107] to evaluate our method. Table 8 shows the experimental results of the different methods. We make a comparison with DeconvNet [109], FCN-8 [95], Ensemble method [107], U-Net [120] and Stacked U-Net. On the TNBC dataset, our model achieved the highest score both in terms of F_1 and AJI compared to other method, again. Compared to the highest obtained by other method, our model approximately 3.1 percent higher in AJI and 1.3 percent higher in F_1. Moreover, for AJI and F_1, all of our three models are perform better than all other methods.

As a whole, these experimental results indicate that our model with two improvements (**Ours (DLA)**) performs significantly better and achieved the highest overall AJI and F_1 scores compared with other segmentation methods both on the TCGA and TNBC datasets. Even the performance of our model with just one improvement (**U-Net (DLA)** or **Ours**) is better than the majority of the other methods.

Specifically, the proposed two-stage learning framework with stacking U-Nets (**Ours**) obtained

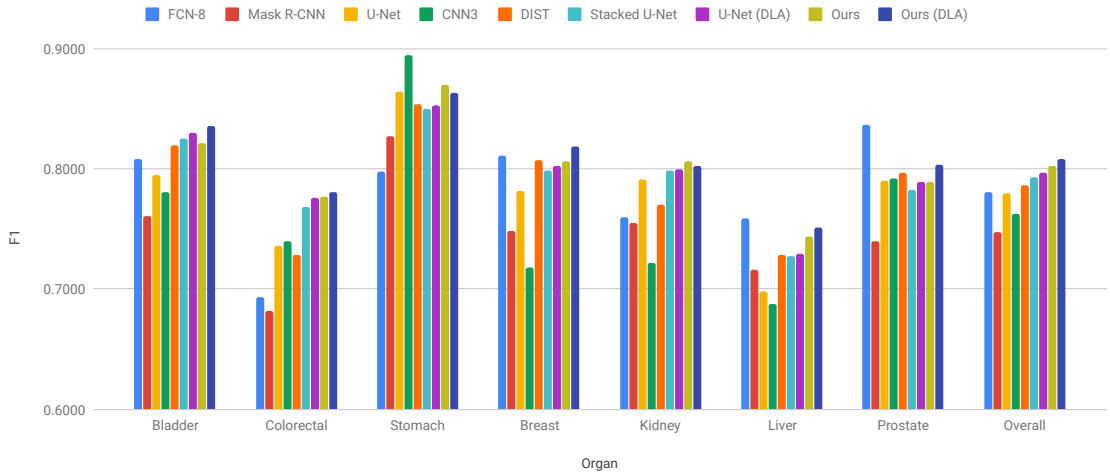

Figure 31: Comparative analysis of F_1 scores for each organs on the TCGA test set.

Table 7: AJI and F_1 scores of different methods on the same organ testing set and different organ testing set of the TCGA dataset.

Method	AJI↑		F_1 Score ↑	
	Same organ testing	Different organ testing	Same organ testing	Different organ testing
FCN-8 [95]	0.5382	0.4891	0.7916	0.7667
Mask R-CNN [56]	0.5010	0.4992	0.7398	0.7566
U-Net [120]	0.5070	0.5331	0.7654	0.7984
CNN3 [81]	0.5154	0.4989	0.7301	0.8051
DIST [108]	0.5594	0.5604	0.7756	0.8005
Stacked U-Net	0.5286	0.5724	0.7770	0.8144
U-Net (DLA)	0.5403	0.5825	0.7802	0.8194
Ours	0.5633	0.5987	0.7865	0.8229
Ours (DLA)	**0.5746**	**0.6094**	**0.7939**	**0.8266**

much more superior performance metrics than the model which just stack two U-Nets without two-stage learning framework (Stacked U-Net) both on these two datasets, therefore demonstrating that the proposed two-stage learning framework can achieve performance improvements dramatically for nuclei segmentation task. On the other hand, **U-Net (DLA)** achieved higher AJI and F_1 scores compared to the original U-Net architecture both on the TCGA and TNBC datasets, it proves that DLA can boost the performance of U-Net in nuclei segmentation via learning better image representations. Finally, our model with these two main improvements (**Ours (DLA)**) surpass all other methods including **Ours** and **U-Net (DLA)** with a large margin, further indicates that our two improvements can combine together effectively and achieves great performance beneficial with the corporation of them.

In addition, the experimental results of same organ testing and different organ testing of the TCGA dataset illustrate that our proposed approach has an excellent generalization ability to images come from different organs which do not appear in the training set.

Table 8: Quantitative comparison of different methods on the TNBC dataset.

Method	Recall↑	Precision↑	F_1 ↑	AJI↑
DeconvNet [109]	0.773	**0.864**	0.805	-
FCN-8 [95]	0.752	0.823	0.763	-
Ensemble [107]	**0.900**	0.741	0.802	-
U-Net [120]	0.800	0.820	0.810	0.578
Stacked U-Net	0.802	0.830	0.816	0.580
U-Net (DLA)	0.812	0.826	0.818	0.586
Ours	0.818	0.824	0.821	0.595
Ours (DLA)	0.833	0.826	**0.829**	**0.611**

4.5.5 Qualitative Analysis

Fig. 32 shows some example segmentation output images of our best model (**Ours (DLA)**). Since our model has two outputs, we show the nuclei-boundary and nuclei output and the nuclei segmentation output. From these results, we can observe that except even though there are still some small imperfections inside the segmentation result, the overall segmentation quality is quite good.

Generalization Ability

In order to compare with some traditional nuclei segmentation methods, we implement two of them, one is OTSU adaptive intensity thresholding algorithm, another one is a method based on watershed segmentation algorithm. Fig. 33 and Fig. 34 show the results of our model compared to the results of these two traditional nuclei segmentation methods on the testing set of TCGA. And Fig. 35 gave some example results of our proposed approach compare with the results of the two conventional nuclei segmentation methods on the TNBC dataset. From these results, we can see that our CNN based approach achieved the best performance than traditional approaches, this also can be reflect on the corresponding AJI and F_1 metrics. The results of the two traditional methods often lead to merged nuclei (under-segmentation), but our model handle these challenging situations properly, i.e. nuclei come from different organs and have different size, shape, appearance and density.

Small Nuclei Segmentation

As mentioned before, one of the key challenges in nuclei segmentation is segmenting nuclei which has extremely small size. Fig. 36 gives a concrete comparison example of segmentation results between our model and U-Net. The nuclei in red rectangles are not detected by U-Net but segmented by our proposed model precisely. This may due to that the addition of nuclei boundaries force the segmentation network to pay more attention on small nuclei and therefore can segment them correctly.

Overlapping and Touching Nuclei Segmentation

Another challenging case in nuclei segmentation is the overlapping and touching nuclei. Fig. 37 demonstrates segmentation examples of our proposed model on this kind of situation. We can

clearly observe that the core idea of our proposed model - the addition of nuclei boundaries can improve the performance of segmenting such nuclei compared to baseline widely used medical image segmentation model (i.e. U-Net).

4.6 Conclusion

In this study, we propose a two-stage learning framework based on stacking two U-Nets with DLA for nuclei segmentation. We convert the binary segmentation task into a two-step task inspired by the idea of curriculum learning. The difficulty of segmenting small, overlapping and touching nuclei directly from histopathological images is addressed by introducing nuclei-boundary prediction as the intermediate step. Furthermore, along with the two-step task, we design a two-stage learning framework by stacking two U-Nets, where the task of each U-Net is different but highly-related and trained simultaneously. Finally, DLA is adopted to extend the skip connections in U-Net to better fuse features across different levels for nuclei segmentation.

The experimental results on two public and diverse nuclei datasets demonstrate that our proposed approach outperforms many standard segmentation architectures and the most recently proposed nuclei segmentation methods and can be easily generalized to different organs, tissue and cell types.

In addition, the segmentation results of our proposed model achieved superior performance quantitatively and qualitatively on some challenges cases such as small, overlapping and touching nuclei compared to traditional nuclei segmentation methods and some CNN-based models like U-Net. It verifies that the addition of nuclei-boundary will improve the performance of CNN-based segmentation model significantly.

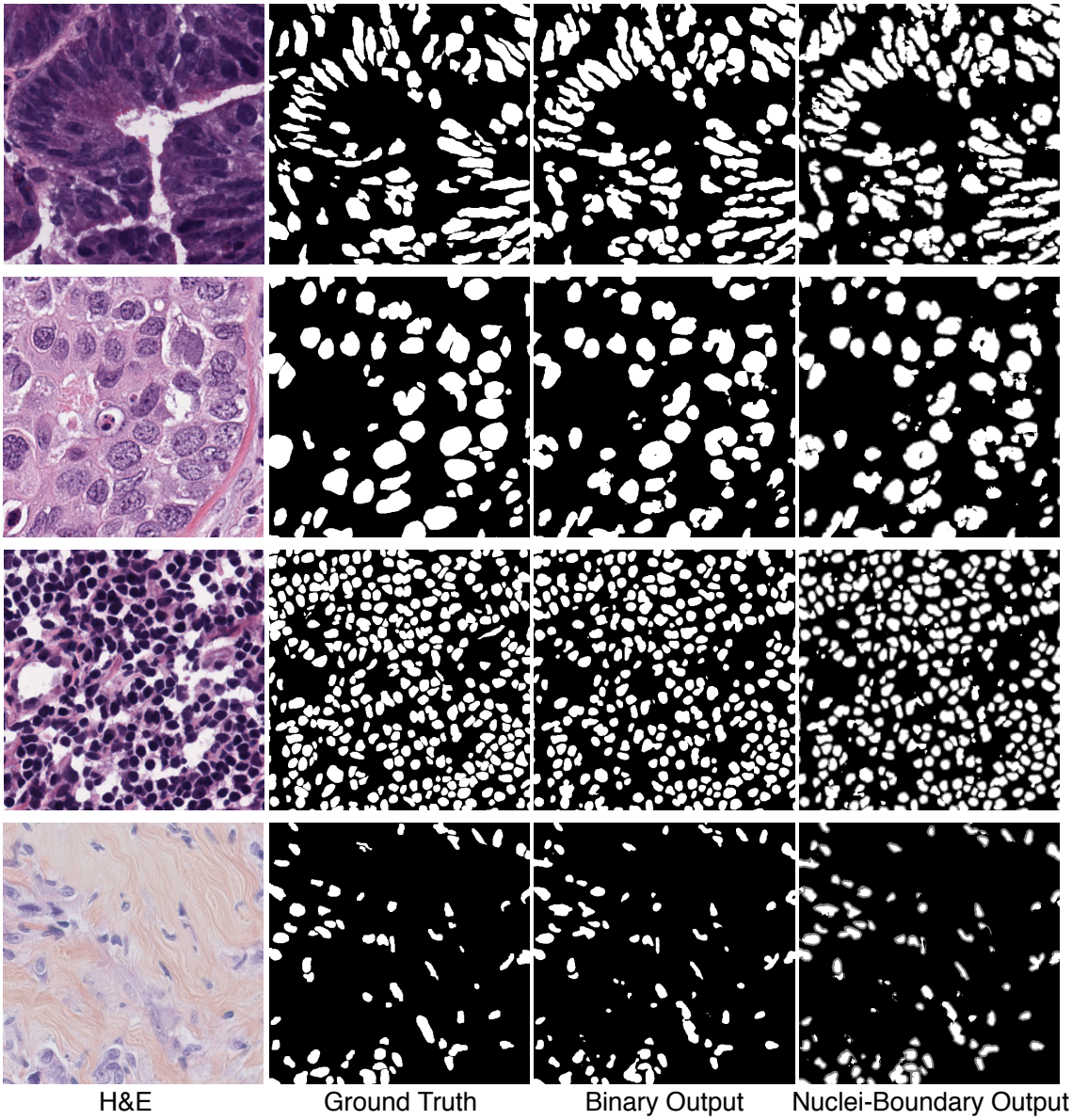

Figure 32: Overall segmentation results. Example input H&E stained images (first column) and associated ground truth (second column) and corresponding binary output (third column) and nuclei-boundary output (forth column). Here we use the outputs of our best model (**Ours (DLA)**). The images of first three rows are from the TCGA dataset and the last one come from the TNBC dataset.

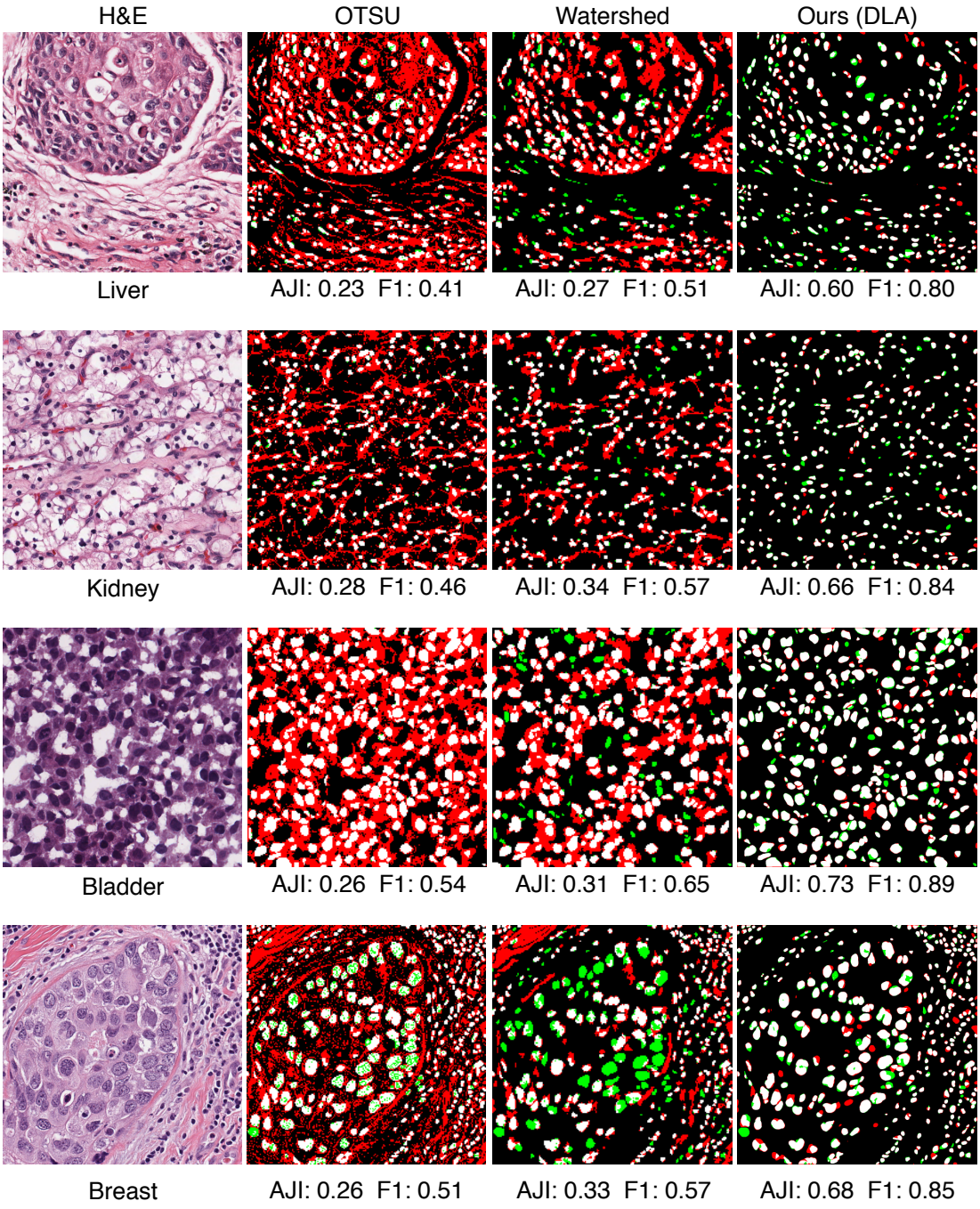

Figure 33: Segmentation results of different methods for different organs (liver, kidney, bladder and breast) on the TCGA dataset. White area indicates True Positives, black area indicates True Negatives, while red area represents False Positive and green area represents False Negative. The associated AJI and F_1 score are shown on the bottom of each result image.

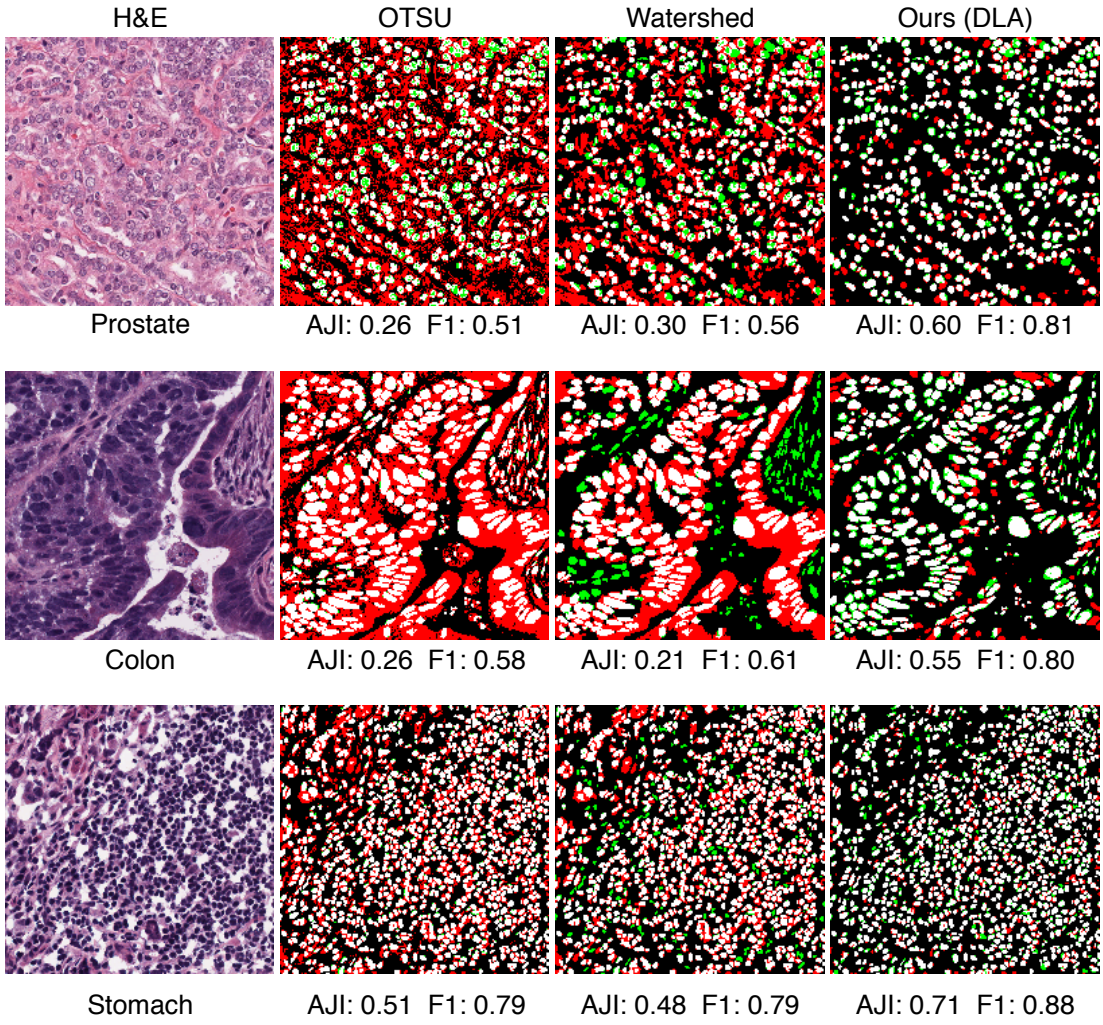

Figure 34: Segmentation results of different methods for different organs (prostate, colon and stomach) on the TCGA dataset. White area indicates True Positives, black area indicates True Negatives, while red area represents False Positive and green area represents False Negative. The associated AJI and F_1 score are shown on the bottom of each result image.

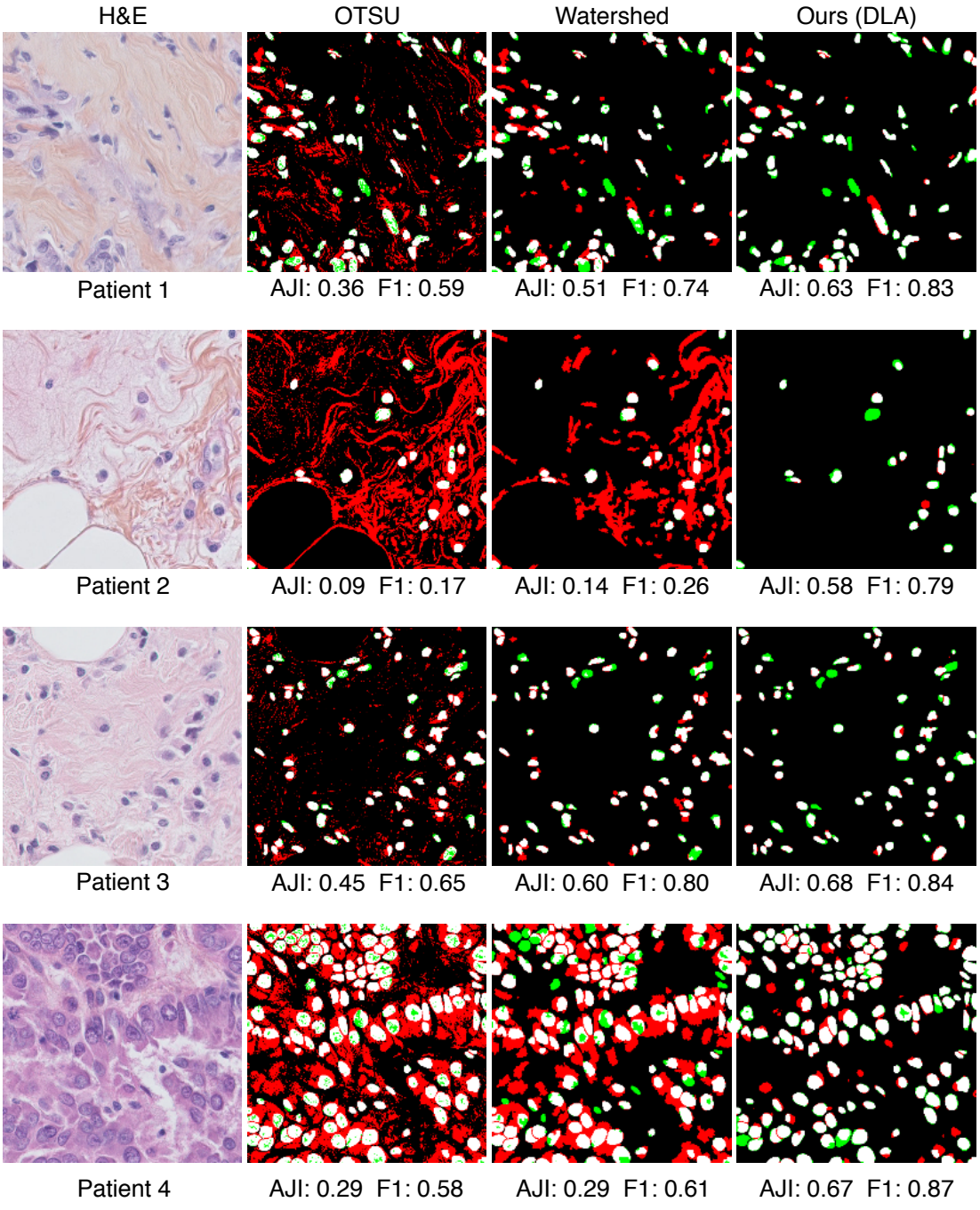

Figure 35: Segmentation results of different methods for different patients on the TNBC dataset. White area indicates True Positives, black area indicates True Negatives, while red area represents False Positive and green area represents False Negative. The associated AJI and F_1 score are shown on the bottom of each result image.

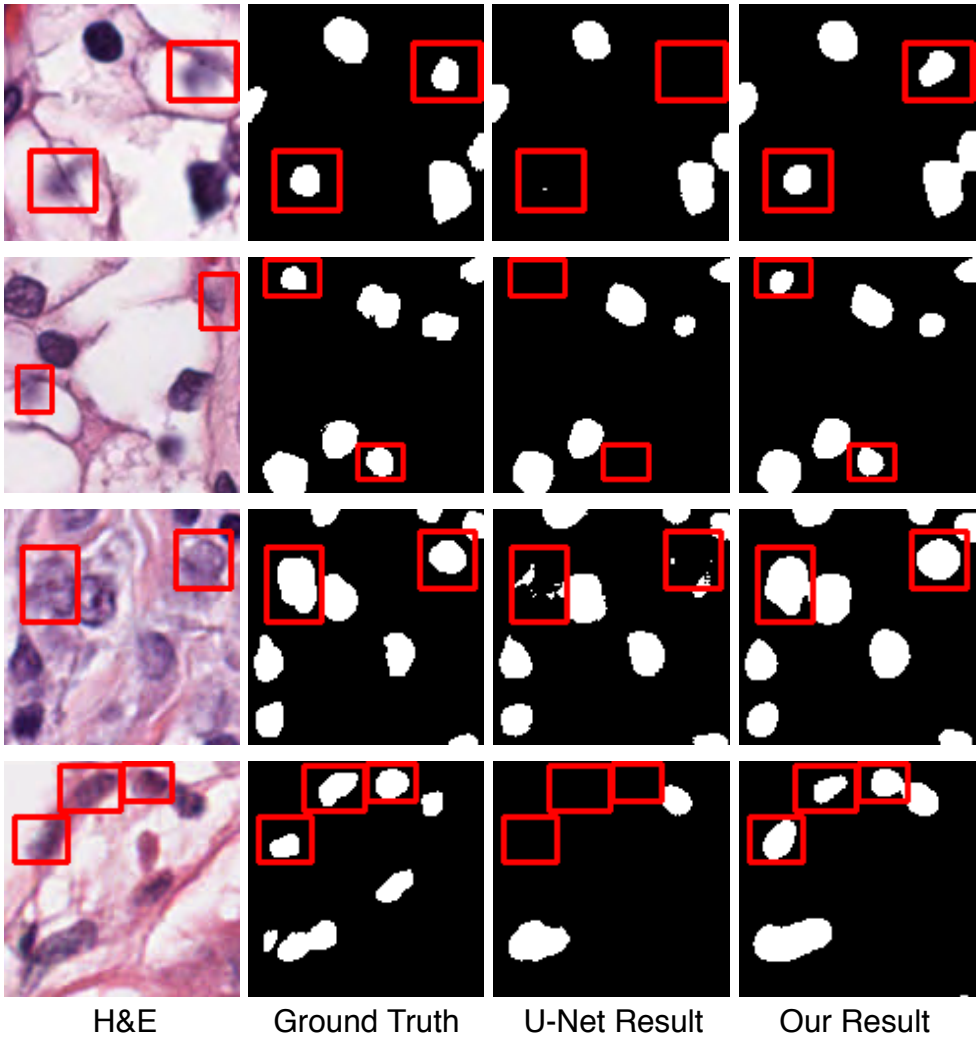

Figure 36: Segmentation results of small nuclei. Example input H&E stained images (first column) and associated ground truth (second column) and corresponding segmentation result of U-Net (third column) and corresponding segmentation result of our model (the last column).

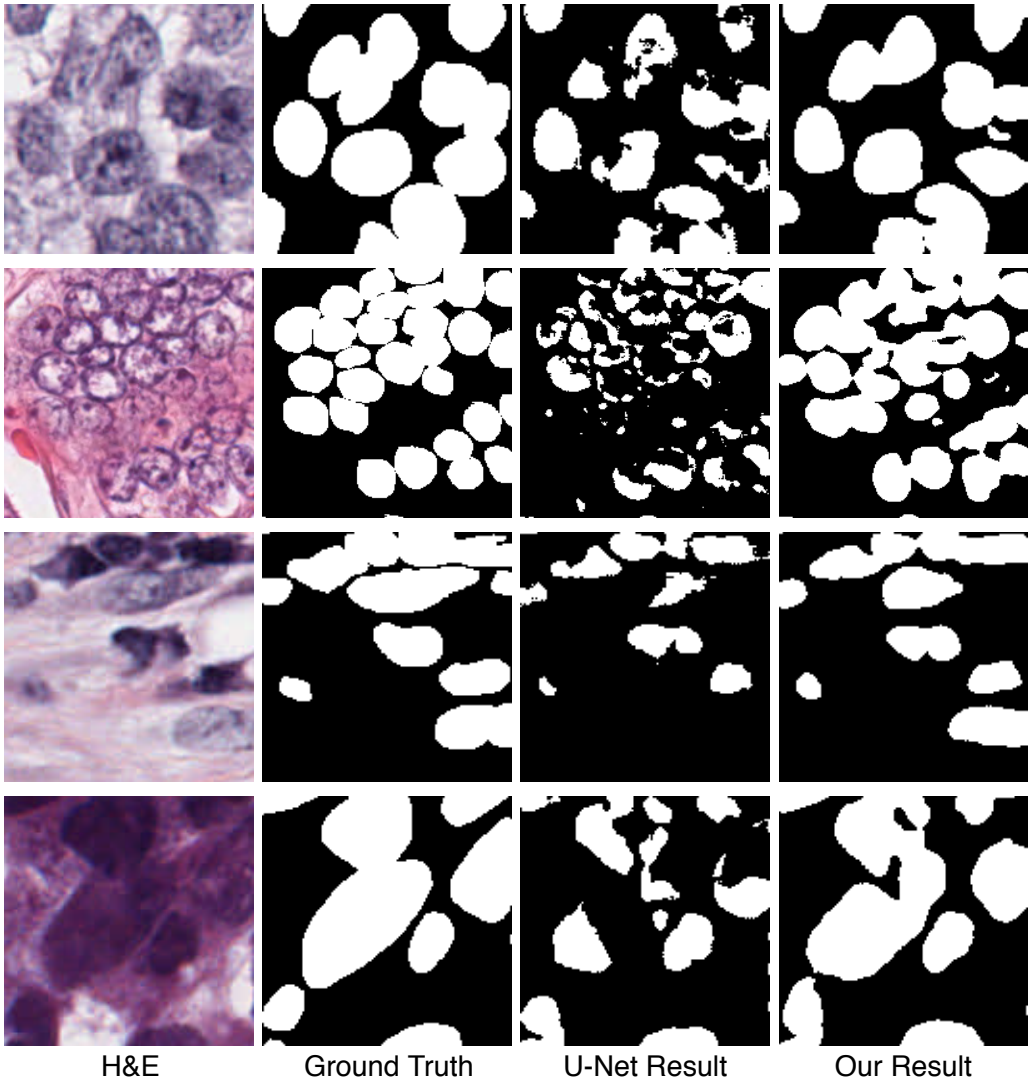

Figure 37: Segmentation results of overlapping and touching nuclei. Example input H&E stained images (first column) and associated ground truth (second column) and corresponding segmentation result of U-Net (third column) and corresponding segmentation result of our model (the last column).

Chapter 5

Conclusion and Future Work

This chapter will give the conclusion and future work of this thesis.

5.1 Conclusion

In this thesis, we focus on CNN based approaches for two different medical image segmentation tasks, i.e. fully convolutional networks for lung segmentation task in chest X-ray images, and two-stage learning framework extends with DLA for nuclei segmentation task in histopathological images.

For the lung segmentation problem, we apply two widely used segmentation model based on fully convolutional networks, i.e. FCN and U-Net, on this task. The experimental results on three publicly available chest X-ray datasets and their combined dataset demonstrate that all CNN based models achieve promising results and the performance of U-Net is the best compared to the FCN models.

For the nuclei segmentation task, since the principle challenge of it is how to segment the small, overlapping and touching nuclei precisely, we propose a two-stage learning framework based on the idea of curriculum leaning. Specifically, we firstly divide the binary segmentation task into a two-step task by introducing the nuclei-boundary prediction as an intermediate step. The addition of boundary areas will force the segmentation network pay more attention on the small nuclei and the overlapping, touching areas between nuclei, also this will decrease the difficulty of segmenting the nuclei directly from input images. To solve the two-step task, we design a two-stage learning framework by cascading two U-Nets, the purpose of the first U-Net is nuclei-boundary prediction while the task of the second U-Net is the prediction of final fine-grained nuclei segmentation map. Furthermore, since the images may come from different medical sites and operated by different physician, and the nuclei have a great diversity in size, shape, appearance and density, in order to increase the generalization ability of our method, we extend the U-Nets with DLA by iteratively merging features across different levels. We adopt two public diverse H&E stained nuclei datasets. The experimental results show that our proposed approach outperforms many standard segmentation architectures and recently proposed nuclei segmentation methods, and can be easily generalized across different cell types in various organs.

5.2 Future Work

In this section, we discuss some possible directions for future research.

5.2.1 More Effective Loss Function for Medical Segmentation Task

In this thesis, only the binary cross entropy loss and the categorical cross entropy loss are used for these two segmentation tasks, one possible research direction is design different loss function suits for different tasks. For example, the loss function for the nuclei segmentation task can give large weight to the pixels inside small nuclei and overlapping or touching areas between nuclei, thus can differentiate these pixels from background pixels.

5.2.2 More Useful Strategies for Training Deep CNNs

The training process is critical for the success of a CNN model, however, this is still be a challenge for deep learning community. Except for the optimization techniques, some other techniques such as deep supervision [87], which the core idea of this technique is to provide additional direct supervision to the hidden layer and propagate it to lower layers instead of just the direct supervision to the output layer, can be used for these two segmentation tasks.

5.2.3 More Deeper and Powerful Networks

For these two research works, we only focus on the application of FCN and U-Net. However, there exist some segmentation works with much deeper networks, for example, the network proposed by [71] has more than 100 layers and dense connections [63], it achieved state-of-the-art performance on urban scene segmentation benchmark. How to apply such a deeper network effectively in the medical image segmentation field remains further research.

Bibliography

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] Wan Siti Halimatul Munirah Wan Ahmad, W Mimi Diyana W Zaki, and Mohammad Faizal Ahmad Fauzi. Lung segmentation on standard and mobile chest radiographs using oriented gaussian derivatives filter. *Biomedical Engineering Online*, 14(1):20, 2015.

[3] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.

[4] Pavan Annangi, Sheshadri Thiruvenkadam, Anand Raja, Hao Xu, XiWen Sun, and Ling Mao. A region based active contour method for x-ray lung segmentation using prior shape and low level features. In *2010 IEEE International Symposium on Biomedical Imaging: from nano to macro*, pages 892–895. IEEE, 2010.

[5] Michela Antonelli, Graziano Frosini, Beatrice Lazzerini, and Francesco Marcelloni. A cad system for lung nodule detection based on an anatomical model and a fuzzy neural network. In *NAFIPS 2006-2006 Annual Meeting of the North American Fuzzy Information Processing Society*, pages 448–453. IEEE, 2006.

[6] Muhammad Arif and Nasir Rajpoot. Classification of potential nuclei in prostate histology images using shape manifold learning. In *2007 International Conference on Machine Vision (ICMV)*, pages 113–118. IEEE, 2007.

[7] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[8] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual International Conference on Machine Learning (ICML)*, pages 41–48. ACM, 2009.

[9] James Bergstra, Frédéric Bastien, Olivier Breuleux, Pascal Lamblin, Razvan Pascanu, Olivier Delalleau, Guillaume Desjardins, David Warde-Farley, Ian Goodfellow, Arnaud Bergeron, et al. Theano: Deep learning on gpus with python. In *NIPS 2011, BigLearning Workshop, Granada, Spain*, volume 3, pages 1–48. Citeseer, 2011.

[10] James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[11] Lei Bi et al. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, 64(9):2065–2074, 2017.

[12] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):1124–1137, 2004.

[13] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 377–384. IEEE, 1999.

[14] Herb Brody. Medical imaging. *Nature*, 502(7473):S81–S81, 2013.

[15] Matthew S Brown, Laurence S Wilson, Bruce D Doust, Robert W Gill, and Changming Sun. Knowledge-based method for segmentation and analysis of lung boundaries in chest x-ray images. *Computerized Medical Imaging and Graphics*, 22(6):463–477, 1998.

[16] G Bueno, R Gonzalez, Oscar Déniz, Marcial García-Rojo, J Gonzalez-Garcia, MM Fernández-Carrobles, Noelia Vállez, and Jesús Salido. A parallel solution for high resolution histological image analysis. *Computer Methods and Programs in Biomedicine*, 108(1):388–401, 2012.

[17] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging*, 33(2):577–590, 2013.

[18] Francisco M Carrascal, José M Carreira, Miguel Souto, Pablo G Tahoces, Lorenzo Gómez, and Juan J Vidal. Automatic calculation of total lung capacity from automatically traced lung boundaries in postero-anterior and lateral digital chest radiographs. *Medical Physics*, 25(7):1118–1131, 1998.

[19] Vicent Caselles, Francine Catté, Tomeu Coll, and Françoise Dibos. A geometric model for active contours in image processing. *Numerische mathematik*, 66(1):1–31, 1993.

[20] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[21] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[22] Jun-Cheng Chen, Vishal M Patel, and Rama Chellappa. Unconstrained face verification using deep cnn features. In *2016 IEEE Winter conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.

[23] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.

[24] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.

[25] Xiao-juan Chen and Dan Li. Medical image segmentation based on threshold svm. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–3. IEEE, 2010.

[26] Kin-Hoe Chow, Rachel E Factor, and Katharine S Ullman. The nuclear envelope environment and its cancer connections. *Nature Reviews Cancer*, 12(3):196, 2012.

[27] Patrick Ferdinand Christ et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *MICCAI*, pages 415–423. Springer, 2016.

[28] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432. Springer, 2016.

[29] Laurent D Cohen. On active contour models and balloons. *CVGIP: Image understanding*, 53(2):211–218, 1991.

[30] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine learning (ICML)*, pages 160–167. ACM, 2008.

[31] Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann Lecun. Very deep convolutional networks for text classification. *arXiv preprint arXiv:1606.01781*, 2016.

[32] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[33] Giuseppe Coppini, Massimo Miniati, Simonetta Monti, Marco Paterni, Riccardo Favilla, and Ezio Maria Ferdeghini. A computer-aided diagnosis approach for emphysema recognition in chest radiography. *Medical Engineering & Physics*, 35(1):63–73, 2013.

[34] A Dawoud. Lung segmentation in chest radiographs by fusing shape information in iterative thresholding. *IET Computer Vision*, 5(3):185–190, 2011.

[35] Ricard Delgado-Gonzalo, Virginie Uhlmann, Daniel Schmitter, and Michael Unser. Snakes on a plane: A perfect snap for bioimage analysis. *IEEE Signal Process. Mag.*, 32(1):41–48, 2015.

[36] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[37] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8599–8603. IEEE, 2013.

[38] Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer, 2016.

[39] Florian Dubost, Gerda Bortsova, Hieab Adams, Arfan Ikram, Wiro J Niessen, Meike Vernooij, and Marleen De Bruijne. Gp-unet: lesion detection from weak labels with a 3d regression network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 214–221. Springer, 2017.

[40] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

[41] Navid Farahani, Anil V Parwani, and Liron Pantanowitz. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol Lab Med Int*, 7:23–33, 2015.

[42] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision (IJCV)*, 59(2):167–181, 2004.

[43] Paweł Filipczuk, Thomas Fevens, Adam Krzyżak, and Roman Monczak. Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE Transactions on Medical Imaging*, 32(12):2169–2178, 2013.

[44] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

[45] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.

[46] Arthita Ghosh, Max Ehrlich, Sohil Shah, Larry Davis, and Rama Chellappa. Stacked u-nets for ground material segmentation in remote sensing imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 257–261, 2018.

[47] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[48] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[49] Dorothy M Greig, Bruce T Porteous, and Allan H Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):271–279, 1989.

[50] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 2019.

[51] Steven Guan, Amir Khan, Siddhartha Sikdar, and Parag Chitnis. Fully dense unet for 2d sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health Informatics*, 2019.

[52] Shengwen Guo and Baowei Fei. A minimal path searching approach for active shape model (asm)-based segmentation of the lung. In *Medical Imaging 2009: Image Processing*, volume 7259, page 72594B. International Society for Optics and Photonics, 2009.

[53] Metin N Gurcan, Laura Boucheron, et al. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147, 2009.

[54] Xiao Han, Chenyang Xu, and Jerry L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):755–768, 2003.

[55] Akira Hasegawa, Shih-Chung Benedict Lo, Matthew T Freedman, and Seong Ki Mun. Convolution neural-network-based detection of lung structures. In *Medical Imaging 1994: Image Processing*, volume 2167, pages 654–662. International Society for Optics and Photonics, 1994.

[56] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2961–2969, 2017.

[57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.

[58] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[59] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision (ECCV)*, pages 630–645. Springer, 2016.

[60] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Brian Kingsbury, et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29, 2012.

[61] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[62] Shiying Hu, Eric A Hoffman, and Joseph M Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. *IEEE Transactions on Medical Imaging*, 20(6):490–498, 2001.

[63] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.

[64] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959.

[65] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.

[66] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968.

[67] Isaac Iglesias, Pablo G Tahoces, Miguel Souto, Anxo Martínez de Alegría, María J Lado, and Juan J Vidal. Lung segmentation on postero-anterior digital chest radiographs using active contours. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 538–546. Springer, 2004.

[68] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, pages 448–456, 2015.

[69] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, 4(6):475, 2014.

[70] Anil K Jain, Jianchang Mao, and KM Mohiuddin. Artificial neural networks: A tutorial. *Computer*, (3):31–44, 1996.

[71] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 11–19, 2017.

[72] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[73] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision (IJCV)*, 1(4):321–331, 1988.

[74] Andrew Khalel and Motaz El-Saban. Automatic pixelwise object labeling for aerial imagery using stacked u-nets. *arXiv preprint arXiv:1803.04953*, 2018.

[75] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[76] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. In *Advances in Neural Information Processing Systems*, pages 6965–6975, 2018.

[77] Hui Kong, Metin Gurcan, and Kamel Belkacem-Boussaid. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. *IEEE Transactions on Medical Imaging*, 30(9):1661–1677, 2011.

[78] Sonal Kothari, Qaiser Chaudry, and May D Wang. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 795–798. IEEE, 2009.

[79] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[80] M Pawan Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems*, pages 1189–1197, 2010.

[81] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017.

[82] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[83] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[84] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[85] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[86] Yann LeCun et al. Generalization and network design strategies. In *Connectionism in Perspective*, volume 19. Citeseer, 1989.

[87] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial Intelligence and Statistics*, pages 562–570, 2015.

[88] Yong Jae Lee and Kristen Grauman. Learning the easy things first: Self-paced visual category discovery. In *CVPR 2011*, pages 1721–1728. IEEE, 2011.

[89] Lihua Li, Yang Zheng, Maria Kallergi, and Robert A Clark. Improved method for automatic identification of lung regions on chest radiographs. *Academic Radiology*, 8(7):629–638, 2001.

[90] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37(12):2663–2674, 2018.

[91] Xuechen Li, Suhuai Luo, Qingmao Hu, Jiaming Li, Dadong Wang, and Fabian Chiong. Automatic lung field segmentation in x-ray radiographs using statistical shape and appearance models. *Journal of Medical Imaging and Health Informatics*, 6(2):338–348, 2016.

[92] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1925–1934, 2017.

[93] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[94] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[95] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[96] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4990–4998, 2017.

[97] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, page 3, 2013.

[98] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[99] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities, 2016.

[100] Ravi Malladi, James A Sethian, and Baba C Vemuri. Shape modeling with front propagation: A level set approach. 1994.

[101] Michael F McNitt-Gray, HK Huang, and James W Sayre. Feature selection in the pattern classification problem of digital chest radiograph segmentation. *IEEE Transactions on Medical Imaging*, 14(3):537–547, 1995.

[102] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[103] Massimo Miniati, Giuseppe Coppini, Simonetta Monti, Matteo Bottai, Marco Paterni, and Ezio Maria Ferdeghini. Computer-aided recognition of emphysema on digital chest radiography. *European Journal of Radiology*, 80(2):e169–e175, 2011.

[104] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.

[105] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

[106] Ebrahim Nasr-Esfahani, Shadrokh Samavi, Nader Karimi, SM Reza Soroushmehr, Kevin Ward, Mohammad H Jafari, Banafsheh Felfeliyan, B Nallamothu, and Kayvan Najarian. Vessel extraction in x-ray angiograms using deep learning. In *2016 38th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 643–646. IEEE, 2016.

[107] Peter Naylor, Marick Lae, Fabien Reyal, and Thomas Walter. Nuclei segmentation in histopathology images using deep neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 933–936. IEEE, 2017.

[108] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019.

[109] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1520–1528, 2015.

[110] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[111] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[112] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[113] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[114] Stefano Pedemonte, Bernardo Bizzo, Stuart Pomerantz, Neil Tenenholtz, Bradley Wright, Mark Walters, Sean Doyle, Adam McCarthy, Renata Rocha De Almeida, Katherine Andriole, et al. Detection and delineation of acute cerebral infarct on dwi using weakly supervised machine learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 81–88. Springer, 2018.

[115] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000.

[116] Ewa Pietka. Lung segmentation in digital radiographs. *Journal of Digital Imaging*, 7(2):79–84, 1994.

[117] Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[118] Adnan Qayyum, Syed Muhammad Anwar, Muhammad Majid, Muhammad Awais, and Majdi Alnowami. Medical image analysis using convolutional neural networks: a review. *arXiv preprint arXiv:1709.02250*, 2017.

[119] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.

[120] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

[121] Artem Sevastopolsky et al. Stack-u-net: Refinement network for image segmentation on the example of optic disc and cup. *arXiv preprint arXiv:1804.11294*, 2018.

[122] Mehmet Sezgin and Bülent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–166, 2004.

[123] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.

[124] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.

[125] Yonghong Shi, Feihu Qi, Zhong Xue, Liya Chen, Kyoko Ito, Hidenori Matsuo, and Dinggang Shen. Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. *IEEE Transactions on Medical Imaging*, 27(4):481–494, 2008.

[126] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000.

[127] S Shostak. Histologys nomenclature: Past, present and future. *Biol Syst Open Access*, 2(122):2, 2013.

[128] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 2015.

[129] Concetto Spampinato, Simone Palazzo, Daniela Giordano, Marco Aldinucci, and Rosalia Leonardi. Deep learning for automated skeletal bone age assessment in x-ray images. *Medical Image Analysis*, 36:41–51, 2017.

[130] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[131] Sergii Stirenko, Yuriy Kochura, Oleg Alienin, Oleksandr Rokovyi, Yuri Gordienko, Peng Gang, and Wei Zeng. Chest x-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In *2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)*, pages 422–428. IEEE, 2018.

[132] James S Supancic and Deva Ramanan. Self-paced learning for long-term tracking. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2379–2386, 2013.

[133] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

[134] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[135] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[136] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[137] Shaghayegh Taheri, Thomas Fevens, and Tien D Bui. Robust nuclei segmentation in cytohistopathological images using statistical level set approach with topology preserving constraint. In *Medical Imaging 2017: Image Processing*, volume 10133, page 1013318. International Society for Optics and Photonics, 2017.

[138] Osamu Tsujii, Matthew T Freedman, and Seong K Mun. Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network. *Medical Physics*, 25(6):998–1007, 1998.

[139] Abhishek Vahadane et al. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, 2016.

[140] Bram Van Ginneken, Alejandro F Frangi, Joes J Staal, Bart M ter Haar Romeny, and Max A Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.

[141] Bram Van Ginneken, BM Ter Haar Romeny, and Max A Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE Transactions on Medical Imaging*, 20(12):1228–1241, 2001.

[142] Bram Van Ginneken, Mikkel B Stegmann, and Marco Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis*, 10(1):19–40, 2006.

[143] Luminita A Vese and Tony F Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[144] Mitko Veta, A Huisman, Max A Viergever, Paul J van Diest, and Josien PW Pluim. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 618–621. IEEE, 2011.

[145] Neal F Vittitoe, Rene Vargas-Voracek, and Carey E Floyd Jr. Identification of lung regions in chest radiographs using markov random field modeling. *Medical Physics*, 25(6):976–985, 1998.

[146] Zhenyu Wu and Richard Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11):1101–1113, 1993.

[147] Fuyong Xing and Lin Yang. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Reviews in Biomedical Engineering*, 9:234–263, 2016.

[148] Chenyang Xu, Jerry L Prince, et al. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[149] Xin-Wei Xu and Kunio Doi. Image feature analysis for computer-aided diagnosis: Detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs. *Medical Physics*, 23(9):1613–1624, 1996.

[150] Jing-Hao Xue and D Michael Titterington. t-tests, f-tests and otsu's methods for image thresholding. *IEEE Transactions on Image Processing*, 20(8):2392–2396, 2011.

[151] Xiaodong Yang, Houqiang Li, and Xiaobo Zhou. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 53(11):2405–2414, 2006.

[152] Xin Yao. Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9):1423–1447, 1999.

[153] Anthony Yezzi, Satyanad Kichenassamy, Arun Kumar, Peter Olver, and Allen Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2):199–209, 1997.

[154] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational intelligenCe magazine*, 13(3):55–75, 2018.

[155] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[156] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. Deep layer aggregation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2403–2412, 2018.

[157] Tianli Yu, Jiebo Luo, Amit Singhal, and Narendra Ahuja. Shape regularized active contour based on dynamic programming for anatomical structure segmentation. In *Medical Imaging 2005: Image Processing*, volume 5747, pages 419–430. International Society for Optics and Photonics, 2005.

[158] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[159] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Robert Fergus. Deconvolutional networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, volume 10, page 7, 2010.

[160] Ziming Zeng, Harry Strange, Chunlei Han, and Reyer Zwiggelaar. Unsupervised cell nuclei segmentation based on morphology and adaptive active contour modelling. In *International Conference Image Analysis and Recognition*, pages 605–612. Springer, 2013.

[161] Ying Zhang, Mohammad Pezeshki, Philémon Brakel, Saizheng Zhang, Cesar Laurent Yoshua Bengio, and Aaron Courville. Towards end-to-end speech recognition with deep convolutional neural networks. *arXiv preprint arXiv:1701.02720*, 2017.

[162] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.